

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 23, 2003, 17:44:27 : Search time 3326 Seconds  
(without alignments)  
3202.534 Million cell updates/sec

Title: US-09-842-364-3  
Perfect score: 366  
Sequence: 1 MASMAVLTWALALSAFSA.....LMEDITHSLDGHSHLGD 366

Scoring table:  
Oligo  
Xgapop 60.0, Xgapext 60.0  
Ygapop 60.0, Ygapext 60.0  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Word size: 6 aa = 18 nt

Total number of hits satisfying chosen parameters: 175711

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODL=frame\_plus\_p2n\_model -DEV=x11  
-O=/cgn2.1/USPTO.spool/US09842364/Runat.16012003.164829.10649/app.query.fasta.1.519  
-DB=genembl -OEMT=fastapi -SUFFIX=olig.rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=olig -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=quality -THR\_MIN=6 -ALIGN=15 -MODE=LOCAL -OUTFMT=plc  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09842364 -RCGN\_1\_1\_2425 -Runat.16012003.164829.10649 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAL -LARGOUBERY -NFC\_SCORES=0 -WAIT -IONGLOC -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -Fgapop=6 -Fgapext=7  
-Ygapop=60 -Ygapext=60 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*  
1: gb\_da.\*  
2: gb\_hlg.\*  
3: gb\_la.\*  
4: gb\_lm.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vt.\*  
15: em\_da.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No. Score Match Length DB ID

← length of match

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	366	100.0	1322	6	AX006773	AX006773 Sequence
2	366	100.0	1349	9	AF202890	AF202890 Homo sapi
3	366	100.0	1867	6	AX006774	AX006774 Sequence
4	366	100.0	1879	6	AX063165	AX063165 Sequence
5	366	100.0	1879	6	AX469877	AX469877 Sequence
6	366	100.0	1889	6	AF202889	AF202889 Homo sapi
7	366	100.0	1894	6	AX061634	AX061634 Sequence
8	312	85.2	5381	6	AX063467	AX063467 Sequence
9	312	85.2	5381	6	AX469879	AX469879 Sequence
10	312	85.2	42479	6	AC074203	AC074203 Homo sapi
11	312	85.2	81001	6	AX063464	AX063464 Sequence
12	312	85.2	81001	6	AX469876	AX469876 Sequence
13	312	85.2	115027	2	AF001480	AF001480 Homo sapi
14	312	85.2	115582	2	AF001481	AF001481 Homo sapi
15	312	85.2	188971	9	AC007707	AC007707 Homo sapi
16	312	76.2	1371	6	AX201372	AX201372 Sequence
17	279	58.2	1334	6	AX073667	AX073667 Sequence
18	213	53.6	181412	2	AC113242	AC113242 Pan trogl
19	196	48.1	592	6	ARI29257	ARI29257 Sequence
20	176	48.1	191656	2	AC118574	AC118574 Lemur cat
21	21	7.1	1341	10	BC011198	BC011198 Mus muscu
22	22	7.1	1819	10	AF327059	AF327059 Mus muscu
23	26	7.1	263744	2	AF401201	AF401201 Mus muscu
24	20	5.5	233619	2	AC118580	AC118580 Oryctolag
25	16	4.4	1282	6	AX006770	AX006770 Sequence
26	16	4.4	1300	10	AF202888	AF202888 Rattus no
27	16	4.4	1834	6	AX006771	AX006771 Sequence
28	16	4.4	1850	10	AF202887	AF202887 Rattus no
29	13	3.6	233619	2	AC118580	AC118580 Oryctolag
30	11	3.0	5377	6	AX323765	AX323765 Sequence
31	11	3.0	165681	2	AC111620	AC111620 Rattus no
32	10	2.7	2098	8	YSC70B3	M28428 S.cerevisia
33	10	2.7	3867	8	YSCPH084	D90346 Saccharomyc
34	10	2.7	12461	8	SC7056	Z49218 S.cerevisia
35	10	2.7	103473	2	AC126952	AC126952 Rattus no
36	10	2.7	138429	8	AF003855	AF003855 Oryza sat
37	10	2.7	177049	8	AF003725	AF003725 Oryza sat
38	9	2.5	199	9	HS187A3R	Z57615 H.sapiens C
39	9	2.5	300	4	S79277	S79277 Chromogranin
40	9	2.5	861	11	AF002080	AF002080 Homo sapi
41	9	2.5	1405	8	AY072011	AY072011 Arabidops
42	9	2.5	2800	6	E12365	E12365 Rat CNK1 cd
43	9	2.5	4274	9	HSMB04274	AL832963 Homo sapi
44	9	2.5	5377	6	AX323766	AX323766 Sequence
45	9	2.5	5573	10	RATRSC2	D50413 Rat Tsc2 mR

RESULT 1

ALIGNMENTS

AX006773  
 LOCUS AX006773 1322 bp DNA linear PAT 06-SEP-2000  
 DEFINITION Sequence 22 from Patent WO0003013.  
 ACCESSION AX006773  
 VERSION AX006773.1 GI:9994814  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 1 (bases 1 to 1322)  
 Leegwater, A.C., Van der Vijet, H.N., Chamuleau, R.A. and Groenink, M.  
 Gene and protein involved in liver regeneration  
 JOURNAL: LEEGWATER ADAM CORNELIS JOZEF (NL); VIJET HENDRIK NIELS V D (NL); AMSTERDAM MOLECULAR THERAPEUTI (NL); CHAMULEAU ROBERT ANTOINE FRANC (NL); GROENINK MARTIN (NL)  
 Location/Qualifiers  
 source 1..1322  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 BASE COUNT 267 a 408 c 427 g 220 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0 length: 1322  
 Score: 366.00 Matches: 366  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0  
 US-09-842-364-3 (1-366) x AX006773 (1-1322)  
 QY 1 Metlaasermelaialaivaaleuthrttrpalaleualauleuserialapheseriala 20  
 Db 10 ATGCCAACATCGCTGCGCTCCTCACCCTGCGCTCTTCCTTCAGCCCTTTCCGCC 69  
 QY 21 ThrGlnaiaatgylsglyPheTrpAspTyrPheSerGlnThrSerGlnAspLysGlyArg 40  
 Db 70 ACCCAGGACGCGAAAGGCTTCTGGAGCTACTTACGCCAGACCGAGGCGGAGACCAAGGCGAG 129  
 QY 41 ValGlnGlnIleHisGlnGlnIleHisMetalaargGluProAlaThrIleLysAspSerIleu 60  
 Db 130 GTGGAGCGATTCATCAGCAGAAAGATGCTGCGAGCCCGCGACCTGTAAGACAGCCTT 189  
 QY 61 GlnGlnAspIleuAsnMetalaAsnLysPheLeuGlnIleLysIleuArgProLeuSerGlyIleu 80  
 Db 190 GAGCAAGACCTCAGCAATATGAACAGATTCTCGCAAAAGCTCAGCCCTGAGTGGAGC 249  
 QY 81 GlnAlaProAlaGlyLeuProGlnAspProValGlyMetArgArgGlnIleuGlnIleu 100  
 Db 250 GAGGCTCTCTGGCTCCACAGAGACCGGCTGGGCGATGCGGCGGCGAGCTCAGAGAGAGTTG 309  
 QY 101 GlnGlnValLysAlaArgLeuGlnProTyrMetalaGlnAlaHisGlnIleuValGlyTyr 120  
 Db 310 GAGGAGGTGAAGGCTCGCTCAGCCCTACATGGCAGAGCGCAGAGAGCTGGTGGGCTGG 369  
 QY 121 AsnLeuGlnGlyLeuArgGlnGlnIleuLysProTyrThrMetalaPleuMetGlnGlnVal 140  
 Db 370 AATTGGAGGCTTCGCGCAGCACTGAAAGCCCTACACCATGATGTATGAGAGAGGTG 429  
 QY 141 AlaLeuArgValGlnGlnIleuGlnGlnIleuArgValGlnGlnIleuAspThrLysAla 160  
 Db 430 GCCCTGCGCGCGAGAGAGCTGAGAGAGAGTTCGCGGTGGGGGAGAGACCAAGGCC 489  
 QY 161 GlnLeuLeuGlnGlyValAspGlnAlaTrpAlaLeuLeuGlnGlnIleuSerArgVal 180  
 Db 490 CAGTTGGCTGGGGGGGTGAGAGAGCTTGGCTGGCAGGAGACTCAGAGCGCGCTG 549  
 QY 181 ValHisHisThrGlyArgPheLysGlnIleuPheHisProTyrAlaGlnSerLeuValSer 200  
 Db 550 GTGCACACACCGCGCGCTTCAGAGAGCTTCACCCATACGCCGAGAGAGCTGGTGAGC 609

QY 201 GilyllecIArgHisValGlnGlnIleuHisArgSerValAlaProHisAlaProAlaSer 220  
 Db 610 GGATCGGGGGCCACGTGACAGAGCTGCACCGCATGTGGCTCCGACGCCCCGCCAGC 669  
 QY 221 ProAlaArgLeuSerArgCysValGlnValLeuSerArgLysIleuThrLeuLysAlaLys 240  
 Db 670 CCGCGCCGCTCAGTCGCTGCGCTGCGAGTCTCTCCGGAAGCTACCTCAGCGCCAAAG 729  
 QY 241 AlaLeuHisAlaArgIleGlnGlnIleuLysAspGlnIleuArgGlnIleuSerArgAla 260  
 Db 730 GCCCTGCACGACCATGATCAGACAGAACTGCAGCAGCTCCGCGAGAGCTCAGCAGACCC 789  
 QY 261 PheAlaGlyThrGlyThrGlnGlnIleuValGlnIleuArgProAlaProGlnIleuSerGln 280  
 Db 790 TTGGAGGACCTGGGAGCTGAGAGAGGGCCGCGCGGACCCGACATGCTCTCCGAGAG 849  
 QY 281 ValArgGlnIleuGlnAlaPheArgGlnAspThrTyrLeuGlnIleAlaAlaPheThr 300  
 Db 850 GTCCGCGACGACCTTCAGAGCTTCCGCGAGAGACCTACCTGACATGAGCTTCACT 909  
 QY 301 ArgAlaIleAspGlnGlnThrGlnGlnValGlnGlnIleuAlaProProProGly 320  
 Db 910 CCGCCATCGACACAGACACTGACAGCTCCACGACAGCTCCGCGACCTCCACAGC 969  
 QY 321 HisSerAlaPheAlaProGlnPheGlnGlnIleuThrAspSerGlyLysValLeuSerLysLeu 340  
 Db 970 CACAGTGCCTTCCGCGCGAGAGTTTCACACAAACAGACAGTGGCAGAGCTTGACGAC 1029  
 QY 341 GlnAlaArgLeuAspAspLeuTrpClnuAspIleThrHisSerLeuHisAspGlnIleHis 360  
 Db 1030 CAGCCCGCTCGATGAGCTGCGGAGAGACATCAGCTCAGACGCTTCATACAGAGCGCCAC 1089  
 QY 361 SerHisLeuGlyAspPro 366  
 Db 1090 ACCCATCTGGGGGACCCC 1107  
 RESULT 2  
 LOCUS AF202890 1349 bp mRNA linear PRI 12-APR-2002  
 DEFINITION Homo sapiens apolipoprotein A5 (APOA5) mRNA, complete cds,  
 alternative transcript.  
 ACCESSION AF202890  
 VERSION AF202890.1 GI:6707434  
 KEYWORDS  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 1 (bases 1 to 1349)  
 Pennacchio, L.A., Olivieri, M., Hubacek, J.A., Cohen, J.C., Cox, D.R., Fruchart, J.C., Krauss, R.M. and Rubin, E.M.  
 An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing  
 Science 294 (5540), 169-173 (2001)  
 JOURNAL  
 MEDLINE 21472646  
 PUBMED 11588264  
 REFERENCE  
 2 (bases 1 to 1349)  
 van der Vijet, H.N., Samuels, M.G., Leegwater, A.C., Levels, J.H., Reitsma, P.H., Boers, W. and Chamuleau, R.A.  
 Apolipoprotein A-V: a novel apolipoprotein associated with an early phase of liver regeneration  
 JOURNAL  
 MEDLINE 21576234  
 PUBMED 11577099  
 REFERENCE  
 3 (bases 1 to 1349)  
 Van der Vijet, H.N., Groenink, M., Leegwater, A.C.J. and Chamuleau, R.A.F.M.  
 Direct Submission  
 TITLE Submitted (09-NOV-1999) Experimental Hepatology, Academic Medical Center, Meibergdreef 9, Amsterdam 1105 AZ, Netherlands  
 FEATURES  
 source 1..1349

gene  
CDS  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue="liver"  
/note="Caucasian"  
1..1349  
/gene="APOA5"  
18..1109  
/gene="APOA5"  
/note="constituent of HDL: concentration elevated after partial hepatectomy; alternative polyadenylation; short transcript"  
/codon\_start=1  
/product="apolipoprotein A5"  
/protein\_id="AA025662.1"  
/db\_xref="GI:6707435"  
/translation="MAAVITWALALASATQAKKFWYPSQSCDKWEQIHQO  
KMARPATLAKSLKODINNNKFLKRLPLSOSFAPRIIPDVGMRKQIQEPLFEVKA  
RLQPMVAEHLGVNLEGLRQLEKPYTMDLMEYALRVQELQEDLRVGEPTKQL  
GVDEAMALLOGLOSRRVHITGRKELFHPYASVSGIRHVOELHRSVAPAS  
ARLSRCVYLSTRKTLKAKALHAIQMDLRELSRAFGTGEEGGPPOMLSE  
FVRHRIQAFHODTYLIQIAPIRAIPDEFEVQOQLAPPHPHSAPAFHQQDSKVL  
SKLARLIDMLPDIITHSHIDQCHSLGDP"

BASE COUNT 295 a 408 c 426 g 220 t  
ORIGIN

Alignment Scores:  
Pred. No.: 0 Length: 1349  
Score: 366.00 Matches: 366  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0

US-09-842-364-3 (1-366) x AF202890 (1-1349)

QY 1 MetAlaSerMetAlaAlaValLeuThrTrpAlaLeuAlaLeuSerAlaPheSerAla 20  
DB 9 ATGGCAACATGCGTGGCTGCTCAGCTGGGCTGCTTCTTACAGCTTTTCGGCC 68  
QY 21 ThrGlnAlaArgLysGlyPheTrpAspTrpPheSerGlnThrSerGlyAspLysGlyArg 40  
DB 69 ACCCGAGCAGCGAAAGGCTTCTGGAGTACTTACAGCCAGACGAGGGGACAAAGGACG 128  
QY 41 ValGlnGlnIleHisGlnGlnLysMetAlaArgGluProAlaThrLeuLysAspSerLeu 60  
DB 129 CTCGACGACATTCATCAGCAGACATGCTCCGACGCCGCCAGCCCTGAAACACACGCTT 188  
QY 61 GlnGlnAspLeuAsnAsnMetLysLysPheLeuGlnLysLeuArgProLeuSerGlySer 80  
DB 189 GAGCAAGACCTCAACATATATGAAAGTTCTGGAAGAGCTGAGGCTCTGTGAGTGGAGC 248  
QY 81 GluAlaProAlaGluProGlnAspProValGlyMetArgArgGlnLeuGlnGlnLeu 100  
DB 249 GAGGCTCTCGGCTCCACAGAGACCGGCTGGCATGCGGCGGACCTGACAGAGAGATTG 308  
QY 101 GluGluValLysAlaArgLeuGlnProTrpMetAlaGlnAlaHisGlnLeuValGlyTrp 120  
DB 309 GAGGAGGTGAAGGCTCCGCTCAGCCCTACATGACAGACGCCACAGCTGCTGGCTGG 368  
QY 121 AsnLeuGluGlyLeuArgGlnGlnLeuLysProTrpThrMetAspLeuMetGlnVal 140  
DB 369 AATTGGAGGGCTTGGCGGACCACTGAAGCCCTACAGATGATGTAGAGCAGGTG 428  
QY 141 AlaLeuArgValGlnGlnLeuGlnGlnLeuArgValValGlyGluAspThrLysAla 160  
DB 429 GCCCTGGCGCGTGCAGAGACTCAGAGACAGTTGGCGGTGGGGGAAAGACCCAAAGGC 488  
QY 161 GlnLeuLeuGlyGlyValAspGlnAlaTrpAlaLeuLeuGlnGlyLeuGlnSerArgVal 180  
DB 489 CAGTTGCTGGGCGGCTGCAGAGAGCTTGGCTTGTGACAGGACCTGACAGCGCGCTG 548  
QY 181 ValHisHisThrGlyArgPheLysGlnLeuPheHisProTrpAlaGlnSerLeuValSer 200

DB 549 GTGCACACACCGCGGCTTCAAAAGACTCTTCCACCCTATCGCCGAGAGCTGTGAGC 608  
QY 201 GlyIleGlyArgHisValGlnGlnLeuHisArgSerValAlaProHisAlaProAlaSer 220  
DB 609 GCGATCGCGCGCCACGCTGCAGAGACTTCACCGCAGTGGCTCCGACACCGCCCGCCAGC 668  
QY 221 ProAlaArgLeuSerArgCysValGlnValLeuSerArgLysLeuThrLeuLysAlaLys 240  
DB 669 CCGCGCGCGCTCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 728  
QY 241 AlaLeuHisAlaArgIleGlnGlnAsnLeuAspGlnLeuArgGlnGlnLeuSerArgAla 260  
DB 729 GCGCTGCAGCGACGATCAGCAGACCAAGACTGCGGAGTGGCGGAAAGACTCAGCAGAGC 788  
QY 261 PheAlaGlyThrGlyThrGlyGlnGlyAlaGlyProAspProGlnMetLeuSerGlnGlu 280  
DB 789 TTTTCACAGCGACCTGCGAGTACAGAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 848  
QY 281 ValArgGlnArgLeuGlnAlaPheArgGlnAspThrTrpLeuGlnIleAlaAlaPheThr 300  
DB 849 GTCCGCGCAGCAGCTTACGCTTTCGCCGACAGACACCTACCTGACATAGCTGCTGCTGCT 908  
QY 301 ArgAlaIleAspGlnGlnThrGlnGlnValGlnGlnGlnLeuAlaProProProProGly 320  
DB 909 CGCGCCATGACAGCAGAGACTGAGAGAGTCCAGAGAGTGGCGCCACTCTACACAGGC 968  
QY 321 HisSerAlaPheAlaProGluPheGlnGlnIleThrAspSerGlyLysValLeuSerLysLeu 340  
DB 969 CACAGTGCCTTTCGCCCGCAGAGTTTCAACAAACAGACAGTGGCAGAGCTTTCAGCAAGCTG 1028  
QY 341 GlnAlaArgLeuAspAspLeuTrpGluAspIleThrHisSerLeuHisAspGlnGlyHis 360  
DB 1029 CAGCGCGCTGATGATGACCTGTCGGAACATCATCATCATCATCATCATCATCATCATCAT 1088  
QY 361 SerHisLeuGlyAspPro 366  
DB 1089 AGCAATCTGGGGGAGCCCG 1106

#### RESULT 3

LOCUS AX006774 1867 bp DNA linear PAR 06-SEP-2000  
DEFINITION Sequence 23 from Patent WO0003013.  
ACCESSION AX006774  
VERSION AX006774.1 GI:9994815  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1867)  
AUTHORS Leegwater, A.C., Van der Vliet, H.N., Chamuleau, R.A. and Groenink, M.  
TITLE Gene and protein involved in liver regeneration  
JOURNAL Patent: WO 0003013-A 23 20-JAN-2000;  
LEEGWATER ADAM CORNELIS JOZEF (NL); VLIET HENDRIK NIELS V D (NL);  
AMSTERDAM MOLECULAR THERAPEUTI (NL); CHAMULEAU ROBERT ANTOINE FRANCOIS (NL); GROENINK MARTIN (NL)

FEATURES  
source location/Qualifiers  
1..1867  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 391 a 539 c 585 g 352 t  
ORIGIN

Alignment Scores:  
Pred. No.: 0 Length: 1867  
Score: 366.00 Matches: 366  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0

US-09-842-364-3 (1-366) x AX006774 (1-1867)

1 MetAlaSerMetAlaValLeuThrTrpAlaLeuAlaLeuSerAlaPheSerAla 20  
10 ATGGCAAGCATGCTCCGCTCCACCTGGCTCTGCTCTTTCTTTCAGCGCTTTTCGCC 69  
21 ThrGlnAlaArgIysGlyPheTrpAspTyrPheSerGlnIntrSerGlyAspIysGlyArg 40  
70 ACCCAGGACGGAAGGCTTCTGGACTACTTCCAGCAGACGAGCGGGGACAAAGGAC 129  
41 ValGlnGlnIleHisGlnGlnIleLysMetAlaArgGluProAlaThrLeuLysAspSerLeu 60  
130 GTGGACACATCATCATCAGCAGAAAGATGCTCCGACGCCGCCGACCTTGAACACCTTT 189  
61 GlnGlnAspLeuAsnAsnMetAlaLysPheLeuGlnLysLeuArgProLeuSerGlySer 80  
190 GACCAAGACCTCAACATATGACCAACTCTCTGAAAACCTGACCCCTCTGACTGGAGC 249  
81 GluAlaProArgLeuProGlnAspProValGlyMetArgArgGlnLeuGlnGlnLeu 100  
250 GAGGCTCTCGGCTCCACAGGACCCGGTGGGCTGCGGGGAGCTGCGAGGAGAGTTG 309  
101 GlnGlnValLysAlaArgLeuGlnProTrpMetAlaGlnAlaHisGlnLeuValGlyTrp 120  
310 GAGCAGGTGAAGGCTCCGCTCCAGCTTACATGCGACGAGCGGACGACCTGCTGGCTGG 369  
121 AsnLeuGlnGlnLeuArgGlnGlnLeuLysProTyrTrpMetAlaLeuMetGlnGlnVal 140  
370 AATTTGGAGGGCTTCGCGGACGACCTGACCTGACCTGATCTGATGACGACGAGTGG 429  
141 AlaLeuArgValGlnGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 160  
430 GCCCTCGGCTGAGGAGCTGCGAGGAGCTGCGGCTGGTGGGGGAGACACCAAGGCC 489  
161 GlnLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 180  
490 CAGTCTGCTGGGGGCTGCGAGGAGCTTGGCTTCTGCGAGGAGCTGCGAGGAGCTG 549  
181 ValHisHisThrGlyArgPheLysGlnLeuPheHisProTyrAlaGlnSerLeuValSer 200  
550 GTCCACACACCGGCGCTTCAAGAGCTTCCACACCTGACCTGACCTGACCTGACCTG 609  
201 GlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 220  
610 GGCATCGGGGCGGCGCTGAGGAGCTGACCGGAGTGGCTGCGGCGGCGGCGGCGGCG 669  
221 ProAlaArgLeuSerArgGlyValGlnValLeuSerArgLysLeuThrLeuLysAlaLys 240  
670 CCGCGCGCGCTCACTGCGCTGCGGAGCTGCTGCGGAGGAGCTGCGGAGGAGCTG 729  
241 AlaLeuHisAlaArgIleGlnGlnIleAsnLeuAspGlnLeuArgGlnLeuSerAlaAla 260  
730 GCCCTGACCGGCGGCTGCGGAGCTGCGGAGCTGCGGAGCTGCGGAGCTGCGGAGCT 789  
261 PheAlaGlnIleThrGlnIleThrGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 849  
790 TTTCAGGAGCAGTGGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 849  
281 ValArgGlnArgLeuGlnAlaPheArgGlnAspThrTyrLeuGlnIleAlaAlaPheThr 300  
850 GTGGCGCGGCGCTCACTGCGCTGCGGAGCTGCTGCGGAGGAGCTGCGGAGGAGCT 320  
301 ArgAlaIleAspGlnGlnIleThrGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 340  
910 CCGCGGCGGCTGCGGAGCTGCGGAGCTGCGGAGCTGCGGAGCTGCGGAGCTGCGGAG 360  
321 HisSerAlaPheAlaProGlnIleThrGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 380  
970 CACAGGAGCTGCGGAGGAGTTCACAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 400  
341 GlnAlaArgLeuAspLeuTrpGlnAspIleThrHisSerLeuHisAspGlnGlnIleHis 420  
1030 CAGGCGCGTCTGAGTCACTGCGGAGGAGCTGCGGAGGAGCTGCGGAGGAGCTGCGGAG 440  
361 SerHisLeuGlnAspPro 366

Db 1090 AGCCATCTGGGGGACCCC 1107  
RESULT 4  
AX063465 1879 bp DNA Linear PAT 24-JAN-2001  
LOCUS Sequence 2 from Patent WO0100803.  
DEFINITION AX063465  
ACCESSION AX063465 GI:12541235  
VERSION  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eultheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1879)  
AUTHORS Yen, F., Denison, B., Bour, B., Hithin, B., Hougueler, L., Duclert, A.  
and dumas mline Edwards, J.B.  
TITILE Apolipoprotein A-IV-related protein: polypeptide, polynucleotide  
sequences and diallelic markers  
JOURNAL Patent: WO 0100803-A 2 04-JAN-2001;  
GENSW (FR)  
FEATURES  
source location/Qualifiers  
1..1879  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
5'UTR  
30..1121  
/note="unnamed protein product"  
CDS  
/codon\_start=1  
/protein\_id="CAC25092.1"  
/db\_xref="GI:12541236"  
/translation="MAAVLTWALALSLAFSTQARKGFMDYFSQTSQDGRVEQIHOQ  
/KAREPATIKDLSIEDLNMNMFLEKRLPSSSEARLQVODYGMKROLOELEYJA  
RLOPYMAEHEHIVGMRLBGLROOLKPYTMDLMOVALVQELQBOIRVAGDPTKQOL  
GVDSEWALLOGLQSRVNVHHTGKELFPHYASLVSGIGRHVOIRNSVAHPAPSP  
ARLSRCVOYLSRKLTLKALNLRIONIDOREELSRFAVGTGTEEGAPDPMQSE  
EVORLOAEFODTYLQIAFTRAIDDETEVOQLAPPPGHSAFAPEFOQIDSGKVL  
SKQARLDLMDITSLIDQGHSLDLP  
1122..1879  
3'UTR  
variation  
/note="17-41-250 : polymorphic base C or T"  
BASE COUNT 395 a 540 c 590 g 353 t 1 others  
ORIGIN  
Alignment Scores:  
Pred. No.: 0 Length: 1879  
Score: 366.00 Matches: 366  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0  
DB:  
US-09-842-364-3 (1-366) x AX063465 (1-1879)  
QY 1 MetAlaSerMetAlaValLeuThrTrpAlaLeuAlaLeuSerAlaPheSerAla 20  
10 ATGGCAAGCATGCTCCGCTCCACCTGGCTCTGCTCTTTCTTTCAGCGCTTTTCGCC 80  
21 ThrGlnAlaArgIysGlyPheTrpAspTyrPheSerGlnIntrSerGlyAspIysGlyArg 40  
70 ACCCAGGACGGAAGGCTTCTGGACTACTTCCAGCAGACGAGCGGGGACAAAGGAC 140  
41 ValGlnGlnIleHisGlnGlnIleLysMetAlaArgGluProAlaThrLeuLysAspSerLeu 60  
141 GTGGACACATCATCATCAGCAGAAAGATGCTCCGACGCCGCCGACCTTGAACACCTTT 200  
61 GlnGlnAspLeuAsnAsnMetAlaLysPheLeuGlnLysLeuArgProLeuSerGlySer 80  
201 GACCAAGACCTCAACATATGACCAACTCTCTGAAAACCTGACCCCTCTGACTGGAGC 260  
81 GluAlaProArgLeuProGlnAspProValGlyMetArgArgGlnLeuGlnGlnLeu 100  
261 GAGGCTCTCGGCTCCACAGGACCCGGTGGGCTGCGGGGAGCTGCGAGGAGAGTTG 320







Db 549 GTGCACACACCGCGGCTTCMAAGAGCTCTCCACCCATACGCCGAGAGCTGTGAGC 608  
 QY 201 GIIYIIEGIYATGHIISValGInGluLeuHISarGSerValAlaProHISAlaProAlaSer 220  
 Db 609 GCGATCGGGCCGACCTTCAGAGCTGACCGACCTGTGGCTCCGACGCCGCCGCCGCCGCC  
 QY 221 ProAlaIarGLeuSerArgCysValGInValLeuSerArgLysLeuThrLeuLysAlaLys 240  
 Db 669 CCGCGCGCCCTCAGTCGCTGCCCTGCAGGTCCCTCCCGAGAGCTCAAGCTCAAGCCCAAG 728  
 QY 241 AlaleuHISAlaArgIIEGInGInasnuLeuAspGInLeuArgGluGluLeuSerArgAla 260  
 Db 729 GCGCTGCACGACGAGATCCAGACAGAACTGGACAGCTGGCGGAGAGAGCTGACGAGAGCC 788  
 QY 261 PheAlaGlyThrGlyThrGluGluGluValAspProAspProGInMetLeuSerGluGlu 280  
 Db 789 TTTTGAGGACCTCGGACGACGAGAAAGGCGCGCGCGACCCCGACATGCTTCGCCAGAGAG 848  
 QY 281 ValArgGInArgLeuGInAlaPheArgGInAspThrTrpLeuGInIIEAlaAlaPheThr 300  
 Db 849 GTGCGCCACCGACCTTCAGAGCTTTCCGCGACGACACCTACATGAGCTTCCTTCACT 908  
 QY 301 ArgAlaIIEasPcInGluThrGluGluValGInGInLeuAlaProProProProGly 320  
 Db 909 CGCGGATCGACGACGAGAGAGTGGAGGAGTCCAGAGAGAGCTGGCGGACCTCCACAGAGC 968  
 QY 321 HISSerAlaPheAlaProGluPheGInGInThrAspSerGlyLysValLeuSerLysLeu 340  
 Db 969 CACAGTGGCTTCGCGCGAGCTTTCACAAACAGACAGCTGCGCAAGCTTCGAGCAAGCTG 1028  
 QY 341 GInAlaIarGLeuAspAspLeuTrpGluAspIIEThrHISerLeuHISarGInGInLys 360  
 Db 1029 CAGGCGCTCTGAGTACCTGTGCGAGACATCACTCAAGCTTCATGACAGAGGCCAC 1088  
 QY 361 SerHISLeuGlyAspPro 366  
 Db 1089 AGCATCTGGGGAGCC 1106  
 RESULT 7  
 AX061634  
 LOCUS AX061634 1894 bp DNA linear PAT 24-JAN-2001  
 DEFINITION Sequence 39 from Patent WO0100806.  
 ACCESSION AX061634  
 VERSION AX061634.1 GI:12406729  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 1894)  
 AUTHORS dumas m, linc Edwards, J. H., Bouguetel, L., and Jobert, S.  
 TITLE Complementary dna's encoding proteins with signal peptides  
 JOURNAL Patient: WO 0100806-A 39 04-JAN-2001;  
 GENSER (FR)  
 FEATURES  
 source  
 location/Qualifiers  
 1..1894  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 21..1121  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAC24995.1"  
 /db\_xref="GI:12406730"  
 /translation="MASMAVLTPMALLSAFSATORAKGFMDYFSQTSQDKHVEQI  
 HOOKMAREPATLKDLSLEODLNNNKKFLKRLPSGSPARLPDPPVGMROJDELEEE  
 VKARKLPYAAEHEIYGMNLEGLRQDLKPTTMDLMEVALRVDLQELRVVGEPTA  
 QILGGVDEARALLQGLQSRVYVHHTGRKELHPRAEIVSGIENHVELRVSAPNAP  
 ASPARLSRCYQVLSRKTLTKAKALHAKIQQNLDLRELSRAAGTGTGEGAGDPDM  
 LSEVERQLOAFQSDYLIQIAFTRAIDQETEEVOQLAPPPCHSAFAFEFOOTDSG  
 KVLSKLOARDLDMEDITHSHDGHSHLDP"  
 21..89  
 /note="Von Heijne matrix score 10.80 seq ALALLSFSATOR/RK"  
 polyA\_signal  
 1858..1863

polyA\_site 1879..1894  
 BASE COUNT 411 a 541 c 588 g 353 t 1 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0  
 Score: 366.00 Length: 1894  
 Percent Similarity: 100.00% Conservative: 366  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0  
 US-09-842-364-3 (1-366) x AX061634 (1-1894)  
 QY 1 MetAlaSerMetAlaAlaValLeuThrTrpAlaLeuAlaLeuSerAlaPheSerAla 20  
 Db 21 ATGCAACCATCGCTCCGCTGCTACCTGCGCTCCGCTCTTCTTTCACCGTTTCCGCC 80  
 QY 21 ThrGInAlaArgLysGlyPheThrAspTrpPheSerGInThrSerGlyAspLysGlyArg 40  
 Db 81 ACCCAGGACCGGAAAGGCTTCGACACTTTCAGCCAGACAGACCGCGCAAGGCAAGCAGC 140  
 QY 41 ValGInGInIIEHISGInGInLysMetAlaIarGluProAlaIarThrLeuLysAspSerLeu 60  
 Db 141 GTGAGACAGATCCATCAGCAGAAAGATGCTCCGAGAGCCCGGACCTGAAAGACAGCTT 200  
 QY 61 GInGInAspLeuAsnAsnMetAsnLysPheLeuGInLysLeuArgProLeuSerGlySer 80  
 Db 201 GAGCAAGACCTTCACAAATATGAACAACTTCGCAAAAGCTGAGCGCTCGAGTGGAGAC 260  
 QY 81 GInAlaIarProArgLeuProGInAspProValGlyMetIarArgGInGInGInGInGIn 100  
 Db 261 CAGGCTCTCGGCTCCACAGAGACCGGCGGACATGCGGCGGACCTTCAGAGAGAGTGG 320  
 QY 101 GInGInValLysAlaIarGLeuGInProTrpMetAlaGInAlaHISGInLeuValGlyTrp 120  
 Db 321 GAGGAGGTGAAGGCTCGGCTCCAGGCTTACATGCGAGAGAGCCAGAGCTGGTGGGCTGG 380  
 QY 121 AsnLeuGInGlyLeuArgGInGInLeuLysProTrpThrMetAspLeuMetGluGInVal 140  
 Db 381 AATTTGAGGCGCTTCGCGGACGACCTGAAGCCCTACACATGATTCATGACGAGAGTGC 440  
 QY 141 AlaleuArgValGInGInLeuGInGInLeuArgValValGInLysAspThrLysAla 160  
 Db 441 GCGCTCGGCTTCAGAGAGCTCCAGACGACGCTGGCGGCTGCGGACACACCAAGGCC 500  
 QY 161 GInLeuLeuGlyLysValAspGInValIarTrpAlaIarLeuGInGInLysSerArgVal 180  
 Db 501 CAGTTGCTGGGGGCGGTGGAGCAGAGCTTGGCTTGTCTGACAGAGACTGACAGAGCGCGTG 560  
 QY 181 ValHISHISThrGlyIarGPhelLysGluLeuPheHISProTrpAlaGluSerLeuValSer 200  
 Db 561 GTGCACACACCGCGCGCTTCMAAGAGCTTCACCCATACGCCGAGAGCTGTGTGAGC 620  
 QY 201 GIIYIIEGIYATGHIISValGInGluLeuHISarGSerValAlaProHISAlaProAlaSer 220  
 Db 621 GCGATCGGGCCGACCTTCAGAGCTGACCGACCTGTGGCTCCGACGCCGCCGCCGCCGCC 680  
 QY 221 ProAlaIarGLeuSerArgCysValGInValLeuSerArgLysLeuThrLeuLysAlaLys 240  
 Db 681 CCGCGCGCCCTCAGTCGCTGCCCTGCAGGTCCCTCCCGAGAGCTCAAGCTCAAGCCCAAG 740  
 QY 241 AlaleuHISAlaArgIIEGInGInasnuLeuAspGInLeuArgGluGluLeuSerArgAla 260  
 Db 741 GCGCTGCACGACGAGATCCAGACAGAACTGGACAGCTGGCGGAGAGAGCTGACGAGAGCC 800  
 QY 261 PheAlaGlyThrGlyThrGluGluGluValAspProAspProGInMetLeuSerGluGlu 280  
 Db 801 TTTTGAGGACCTCGGACGACGAGAAAGGCGCGCGGACCCCGACATGCTTCGAGAGAG 860  
 QY 281 ValArgGInArgLeuGInAlaPheArgGInAspThrTrpLeuGInIIEAlaAlaPheThr 300  
 Db 861 GTGCGCCACCGACCTTCAGAGCTTTCGCGACGACACCTCACTGAGATAGTGCCTTCACT 920

BASE COUNT	1104 a	1434 c	1709 g	1132 t	2 others
ORIGIN					
Alignment Scores:					
Pred. No.:	0				5381
Score:	312.00				312
Percent Similarity:	100.00%				Conservative: 0
Best Local Similarity:	100.00%				Mismatches: 0
Query Match:	85.25%				Indels: 0
DB:	6				Gaps: 0
US-09-842-364-3 (1-366) x AX063467 (1-5381)					
QY	55	ThrlleuLysAspSerLeuGluGlnAspIleuAsnMetAsnLysPheLeuGluLysIleu	74		
DB	2244	ACCCTGAAAGACAGCCCTTGAGCAAGACCTTCMACAAATGTGACAAGTTCCGGAAAGAGTG	2303		
QY	75	ArgProLeuSerGlySerGluAlaIleProArgLeuProGlnAspProValGlyMetArg	94		
DB	2304	AGCCCTCGAATGGCAGCGAGCCGCTCCCTCCCTCCACAGACGCCCGCTGGCCATGCCGGCG	2363		
QY	95	GlnLeuGlnGluGluLeuGluGlnValLysAlaArgLeuGlnProTyrPheValGluAla	114		
DB	2364	CAGCTGCAGGAGACCTTGGACAGCTGGAAGCCCTCCCTCCAGCCCTACATGCCAGAGCG	2423		
QY	115	HisGluLeuValGlyTyrAsnLeuGluGlyLeuArgGlnGlnIleuLysProTyrThrMet	134		
DB	2424	CACGAGCTGTGGGTGGAAATTTGGAGGCTTGGCGGACAACTGAAGCCCTACACGATG	2483		
QY	135	AspLeuMetGluGlnValAlaIleuArgValGlnGluLeuGlnGluIleuArgValAl	154		
DB	2484	GATCTCATCGACACAGGTGGCCCTCCCTCCCTCCGACAGACCTCAGACAGACAGTGGCCGTGG	2543		
QY	155	GlyGluAsnThrLysLagIleuLeuGlyGlyValAspGluAlaIleuLeuGln	174		
DB	2544	GGGGAAACACCAAGGCCCGCTTGGTGGCCGCTCCGACAGAGCTTGGCTTGGCTGGCAG	2603		
QY	175	GlyLeuGlnSerArgValValHisHisThrGlyArgPheLysGluLeuPheHisProTyr	194		
DB	2604	GGATGTGAGAGCCCGTGGTGACACACAGGCCGCTTCAAGAGCTCTCCACCCATAC	2663		
QY	195	AlaGluSerLeuValSerGlyIleGlyArgHisValGlnGluLeuHisArgSerValAla	214		
DB	2664	GCCGAGACCTGGTGATGCGCATCGCCGCCAGCTGCGAGAGCTGCACCGAGTGTGCT	2723		
QY	215	ProHisAlaProAlaSerProAlaArgLeuSerArgCysValGlnValLeuSerArgLys	234		
DB	2724	CCGCAAGCCCGCCAGCCCGCCCGCCCGCCCACTCCCTCGTCCAGCTCTCCCGCGAAG	2783		
QY	235	IleuThrLeuLysAlaLysAlaLeuHisAlaArgGlnGlnAsnLeuAspGlnLeuArg	254		
DB	2784	CTTCAGCTCAAGGCCCAAGGCCCTGCAAGCAGCATTCACACAAACCTGCACAGTGGCG	2843		
QY	255	GluGluLeuSerArgAlaPheAlaGlyThrGlyThrGluGluValGlyProAspPro	274		
DB	2844	GAAGACCTCAGCAGACCTTTCGAGCGACATCGACATCGACAAAGGGCCCGCCCGAAGCC	2903		
QY	275	GlnMetLeuSerGluGluValArgGlnArgLeuGlnAlaPheArgGlnAspThrTyrLeu	294		
DB	2904	CAGATGCTCCCGAGAGAGTGGCCGACCACTTCAGGCTTCCGCGAGCAGACACTCCCTG	2963		
QY	295	GlnIleuAlaPheThrArgAlaIleAspGlnGluThrGluGluValGlnGlnGlnIleu	314		
DB	2964	CAGATAGCTGCTTCACTCCGCGCATGACAGCAGAGACTGGAGGGTCCAGCAGACCTG	3023		
QY	315	AlaProProProGlyHisSerAlaPheAlaProGluPheGlnGlnInThrAspSerGly	334		
DB	3024	GGCCACCTCCACGAGCCGACAGTGCCTTCCCGCCAGAGTTTCAACAACAGCAGATGGCC	3083		
QY	335	LysValLeuSerLysLeuGlnAlaArgLeuAspAspLeuTyrGluAspIleThrHisSer	354		
DB	3084	AAGGTTCTGAGCAAGCTGAGGCCGCTGTGATGACGTCGTGGAGAGACATACTACACAGC	3143		



AUTHORS Yang, L., Hu, P., Wu, H., Tapia-Paez, I. and Roe, B. A.  
 TITLE Homo sapiens Chromosome 22q11 Cosmid cos4  
 JOURNAL Unpublished  
 REMARK This clone is a chimeric cosmid from a patient with t(11;22) translocation and it is from the derivative chromosome 11. There are sequences from both chromosome 11 and 22 in this cosmid 2 (bases 1 to 42479)

REFERENCE: Yang, L., Hu, P., Wu, H., Tapia-Paez, I. and Roe, B. A.  
 TITLE Direct Submission  
 JOURNAL Submitted (18-JUL-2000) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA  
 3 (bases 1 to 42479)  
 Yang, L., Hu, P., Wu, H., Tapia-Paez, I. and Roe, B. A.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-AUG-2000) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA  
 4 (bases 1 to 42479)  
 Yang, L., Hu, P., Wu, H., Tapia-Paez, I. and Roe, B. A.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-OCT-2000) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA  
 On Aug 1, 2000 this sequence version replaced g1:9441829.

COMMENT On Aug 1, 2000 this sequence version replaced g1:9441829.  
 FEATURES  
 source location/Qualifiers  
 1. 42479  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="22q11"  
 /clone="cos4"

HASK COUNT 11501 a 9314 c 9563 g 12101 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0 length: 42479  
 Score: 312.00 Matches: 312  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 85.25% Indels: 0  
 DB: 9 Gaps: 0

US-09-842-364-3 (1-366) x AC074203 (1-42479)

QY 55 ThrLeuLysASPserLeuGluGlnAspLeuAsnAsnMetLysPheLeuGluLysLeu 74  
 DB 33380 ACCCTGAAGAGACGCTTGAAGACGCTCAACATATACAAAGTCTCGAAGAGCTG 33439

QY 75 ArgProIeuSerGlySerGluAlaProArgIeuProGlnAspProValClyMetArg 94  
 DB 33440 AGGCTCTAGTGGAGGAGGAGGCTCTCGGCTCCACAGAGCCGGTGGCATGGCGG 33499

QY 95 GlnLeuGlnGluGluLeuGluGluValLysAlaArgLeuGlnProTyrMetAlaGluAla 114  
 DB 33500 CAGCTGACGAGAGCTTGGAGAGGCTGAAGGCTCCCTCCACCCCTACATGGCAGCGC 33559

QY 115 HisGluLeuValAlaGlyTyrPasnLeuGluGlyLeuArgGlnGlnLeuLysProTyrMet 134  
 DB 33560 CACGAGCTGGTGGCTGGAATTGGAGGCTTGGCGGCAACCTGAAGCCCTACACGATG 33619

QY 135 AspLeuMetGluGlnValAlaLeuArgValGlnGluLeuGlnGluGlnLeuArgValAla 154  
 DB 33620 GATCTGATGGAGCAGAGTGGCTCGCGTGGAGAGCTGACAGACAGATTGGCGTGGT 33679

QY 155 GlyGluAspThrLysAlaGlnLeuLeuGlyGlyValAspGluAlaTyrAlaLeuGln 174  
 DB 33680 GCGGAGACACCAAGCCCAAGTCTGGGGGCGGTGGAGAGCGGCTTGGCTGGCAG 33739

QY 175 GlyLeuGlnSerArgValAlaHisThrGlyArgPheLysGlyLeuPheHisProTyr 194  
 DB 33740 GGACTGCAGAGCGCGGTGGTGCACACCGCGCTTCAAGAGCTCTTCACCCATTAC 33799

QY 195 AlaGluSerLeuValSerGlyIleGlyArgHisValGlnGluLeuHisArgSerValAla 214

DB 33800 GCCAGACGCTCGAGAGCGGATCGGCGCCACCTGCAGAGAGCCACCGACATCGCT 33859

QY 215 ProHisAlaProAlaSerProAlaArgLeuSerArgGlyValGlnValLeuSerArgLys 234  
 DB 33860 CCCAGCCCCCGCCAGCCCCCGCCCTCAGTGCCTGCGTGCAGAGCTCTCCCGGAG 33919

QY 235 LeuThrLeuLysAlaLysAlaLeuHisAlaArgGlnGlnGlnAsnLeuAspGlnLeuArg 254  
 DB 33920 CTCACGCTCAAGGCGAGGCGCTTGCAGCGCATTCAGAGAACTGGACCGAGCTGCG 33979

QY 255 GluGluLeuSerArgAlaPheAlaGlyThrGlyThrGluGluGlyAlaGlyProAspPro 274  
 DB 33980 GAAAGCTTCAGACAGGCTTTCAGAGCATTCAGAGAGCGGCGCGCCGAGACCC 34039

QY 275 GlnMetLeuSerGluGluValArgGlnArgLeuGlnAlaPheArgGlnAspThrTyrLeu 294  
 DB 34040 CACATGCTCTCCAGAGAGTCCGCCACAGCATTCAGGCTTCCGCCAGCACCTACCTG 34099

QY 295 GlnIleAlaAlaPheThrArgAlaIleAspGlnGluThrGluGluValGlnGlnGlnLeu 314  
 DB 34100 CAGATAGCTGCTTCACTCGCGCCATTCAGACAGAGACTGAGAGGTCCAGCAGAGCTG 34159

QY 315 AlaProProProProGlnHisSerAlaPheAlaProGluPheGlnGlnThrAspSerGly 334  
 DB 34160 GCGCCACCTCCACAGAGCCACAGTCCCTTCCGCCACAGATTTCACAAACAGACAGCTGC 34219

QY 335 LysValLeuSerLysLeuGlnAlaArgLeuAspAsnLeuTrpGluAspIleThrHisSer 354  
 DB 34220 AAGCTTCTGAGCAAGCGCGAGCGCCCTGATGACCTGTGGGAAACATCTCACAGC 34279

QY 355 LeuHisAspGlnGlyHisSerHisLeuGlyAspPro 366  
 DB 34280 CTTCATGACCGAGGCGCACAGCCATCTGGGGGAGCCCC 34315

RESULT 11  
 AX063464  
 LOCUS AX063464 81001 bp DNA linear PAT 24-JAN-2001  
 DEFINITION Sequence 1 from Patent WO0100803.  
 ACCESSION AX063464  
 VERSION AX063464.1 GI:12541234  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 Yeh, F., Denison, R., Bour, R., Bihain, B., Bouquelerec, L., Duclert, A.  
 and dumas mine Edwards, J.H.  
 TITLE Apolipoprotein a-1-related protein: polypeptide, polynucleotide  
 sequences and diallelic markers thereof  
 JOURNAL Patent: WO 0100803-A1 04-JAN-2001;  
 GENSET (FR)  
 FEATURES  
 location/Qualifiers  
 1. 81001  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 primer\_bind  
 929..949  
 /note="20-828..pu"  
 primer\_bind  
 1220..1238  
 /note="20-828-311.mis"  
 misc\_feature  
 1227..1251  
 /note="20-828-311.probe"  
 variation  
 1239  
 /note="20-828-311 : polymorphic base C or T"  
 primer\_bind  
 1240..1258  
 /note="20-828-311.mis complement"  
 primer\_bind  
 1357..1377  
 /note="20-828..rp complement"  
 misc\_feature  
 10946..12946  
 /note="5'regulatory region"  
 primer\_bind  
 12029..12050  
 /note="17-42..pu"

primer\_bind 12328..12346  
 /note="17-42-319.mis"  
 misc\_feature 12335..12359  
 /note="17-42-319.probe"  
 variation 12347  
 /note="17-42-319 : polymorphic base C or T"  
 primer\_bind 12348..12366  
 /note="17-42-319.mis complement"  
 primer\_bind 12581..12603  
 /note="17-42.rp complement"  
 exon 12947..12958  
 /note="exon 1"  
 exon 13470..13526  
 /note="exon 2"  
 exon 13641..13752  
 /note="exon 3"  
 exon 14271..15968  
 /note="exon 4"  
 primer\_bind 14992..15012  
 /note="17-41.pu"  
 primer\_bind 15222..15240  
 /note="17-41-250.mis"  
 misc\_feature 15229..15253  
 /note="17-41-250.probe"  
 variation 15241  
 /note="17-41-250 : polymorphic base C or T"  
 primer\_bind 15242..15260  
 /note="17-41-250.mis complement"  
 primer\_bind 15460..15482  
 /note="17-41.rp complement"  
 misc\_feature 15969..17969  
 /note="3' regulatory region"  
 primer\_bind 42070..42090  
 /note="20-841.pu"  
 primer\_bind 42199..42217  
 /note="20-841-149.mis"  
 misc\_feature 42206..42230  
 /note="20-841-149.probe"  
 variation 42218  
 /note="20-841-149 : polymorphic base A or G"  
 primer\_bind 42219..42237  
 /note="20-841-149.mis complement"  
 primer\_bind 42572..42591  
 /note="20-841.rp complement"  
 primer\_bind 45328..45347  
 /note="20-842.pu"  
 primer\_bind 45423..45441  
 /note="20-842-115.mis"  
 misc\_feature 45430..45454  
 /note="20-842-115.probe"  
 variation 45442  
 /note="20-842-115 : polymorphic base A or G"  
 primer\_bind 45443..45461  
 /note="20-842-115.mis complement"  
 primer\_bind 45863..45883  
 /note="20-842.rp complement"  
 primer\_bind 76644..76664  
 /note="20-853.pu"  
 primer\_bind 77039..77057  
 /note="20-853-415.mis"  
 misc\_feature 77046..77070  
 /note="20-853-415.probe"  
 variation 77058  
 /note="20-853-415 : polymorphic base C or T"  
 primer\_bind 77059..77077  
 /note="20-853-415.mis complement"  
 primer\_bind 77166..77185  
 /note="20-853.rp complement"  
 BASE COUNT 20291 a 17705 c 19080 g 23919 t 6 others  
 ORIGIN  
 Alignment Scores: 0 Length: 81001  
 Pred. No.:

Score: 312.00 Matches: 312  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 85.25% Indels: 0  
 DB: Gaps: 0  
 US-09-842-364-3 (1-366) x AX063464 (1-81001)  
 QY 55 ThrLeuLysAspSerLeuGluGlnAspLeuAsnMetAsnLysPheLeuGluLysLeu 74  
 Db 14272 ACCCTGAAAGACAGCCTTGAGCAGACCTCAACATATGAACAGTTCTGGAAAGAGCTG 14331  
 QY 75 ArgProLeuSerGlySerGluAlaProArgLeuProGlnAspProValGlyMetArgArg 94  
 Db 14332 AGGCTCTGAGTGGAGCGAGAGCTCTCTGCTCCACAGAGACCGGTGGGTAGCGGCGG 14391  
 QY 95 GlnLeuGlnGluGluLeuGluGluValLysAlaArgLeuGlnProTyrMetAlaGluAla 114  
 Db 14392 CAGCTGCAGCAGAGAGTGTGAGAGAGTGAAGCTGCGCTCCAGCCCTACATGCGAGAGCGC 14451  
 QY 115 HisGluLeuValGlyTyrAsnLeuGluGlyLeuArgGlnGlnLeuLysProTyrThrMet 134  
 Db 14452 CACGAGCTGGTGGCTGGAAATTTGGAGGGCTTGGCGGACAGAACTGAACCTTACACGATG 14511  
 QY 135 AspLeuMetGluGlnValAlaLeuArgValGlnGluLeuGlnGluGlnLeuArgValAla 154  
 Db 14512 GATCTGATGAGAGCAGAGTGGCCCTGGCGTGCAGAGAGTGCAGAGAGCATTTGGCGGTGGT 14571  
 QY 155 GlyGluAspThrLysAlaGlnLeuLeuGlyGlyValAspGluAlaTrrAlaLeuLeuGln 174  
 Db 14572 GCGGAAAGACACCAAGCCGACGCTGCTGGCGGCGGCGGACGAGAGCTTGGCTTCTCAG 14631  
 QY 175 GlyLeuGlnSerArgValAlaHisHisThrGlyArgPheLysGluLeuPheHisProTyr 194  
 Db 14632 GGACTGCAGAGCCGCGGTGGTGCACACACCGGCCCTTCAAGAGACTTCCACCCATAC 14691  
 Db 14692 GCCGAGAGCCTGTGTAGGCGGATCGCGGCGGCGGCGGAGAGCTGCACCGCAGTGTGCT 14751  
 QY 215 ProHisAlaProAlaSerProAlaArgLeuSerArgCysValGlnValLeuSerArgCys 234  
 Db 14752 CCGCAGCCGCCGCCGAGCCGCCGCGGCTCAGCTCCTCGTCCAGCTCTCCCGCAG 14811  
 QY 235 LeuThrLeuLysAlaLysAlaLeuHisAlaArgGlnGlnGlnAsnLeuAspGlnLeuArg 254  
 Db 14812 CTCACGCTCAAGGCCAGAGCCCTGTCAGCAGCATCCAGAGAACCTTGACAGCTGGCG 14871  
 QY 255 ClnGluLeuSerArgAlaPheAlaGlyThrGlnGluGlyAlaGlyProAspPro 274  
 Db 14872 GAAGAGCTCAGCAGAGCCTTTCAGAGCAGTGGAGTGGAGAGGGCGCGCGGACGCC 14931  
 QY 275 GlnMetLeuSerGluGlnValArgGlnArgLeuGlnAlaPheArgGlnAspThrTyrLeu 294  
 Db 14932 CAGATGCTCTCGAGAGAGTGCCTCCAGGACTTACAGCTTTCGCGAGAGACTACTCTG 14991  
 QY 295 GlnIleAlaIlePheThrArgAlaAlaLeuAspGlnGluGluValGlnGlnGlnLeu 314  
 Db 14992 CAGATAGCTGCTTCACTACCTGGGCGCATCCAGCAGAGAGTGAAGAGTCCAGCAGACTG 15051  
 QY 315 AlaProProProProGlnHisSerAlaPheAlaPheAlaProGluPheGlnGlnHisSerCys 334  
 Db 15052 GCGCCACTCCACAGGCGGCGGAGTGCCTTCCGCGAGGTTTCAACAAAGAGAGAGTGGC 15111  
 QY 335 LysValLeuSerLysLeuGlnAlaArgLeuAspAspLeuTrpGluAspIleThrHisSer 354  
 Db 15112 AAGGTTCGACCAAGCTCAGAGCCGCTGTGATGACCTGTGGAAAGACATACACACAC 15171  
 QY 355 LeuHisAspGlnGlnHisSerHisLeuGlnLysPro 366  
 Db 15172 CTTCATGACAGGAGGCGACAGCAGCATCTGGGAGACCC 15207  
 RESULT 12



AX469876 AX469876 81001 bp DNA linear PAT 09-AUG-2002  
 LOCUS Sequence 1 from Patent WO02053734.  
 DEFINITION AX469876  
 ACCSSION AX469876  
 VERSION AX469876.1 GI:22205149  
 KEYWORDS human.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1  
 AUTHORS Bouqueret, L., Duclet, A., Clusel, C., Dumas, M. E., Yen-Polin, F.,  
 Denison, B., Bihain, B., Bour, B., Ebbers-Reed, D. and Satter-Cid, L.  
 TITLE Methods and compositions for inhibiting neoplastic cell growth  
 JOURNAL Patent: WO 02053734-A 1 11-JUL-2002;  
 GENSET SA (FR)  
 FEATURES  
 source location/Qualifiers  
 1..81001  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 primer\_bind 929..949  
 /note="20-828..pu"  
 primer\_bind 1220..1238  
 /note="20-828-311.mis"  
 misc\_binding 1227..1251  
 /bound\_moiety="20-828-311.probe"  
 variation 1239  
 /note="20-828-311 : polymorphic base C or T"  
 primer\_bind 1240..1258  
 /note="20-828-311.mis complement"  
 primer\_bind 1357..1377  
 /note="20-828..rp complement"  
 misc\_feature 10946..12946  
 /note="5;regulatory region"  
 primer\_bind 12029..12050  
 /note="17-42..pu"  
 primer\_bind 12328..12346  
 /note="17-42-319.mis"  
 misc\_binding 12335..12359  
 /bound\_moiety="17-42-319.probe"  
 variation 12347  
 /note="17-42-319 : polymorphic base C or T"  
 primer\_bind 12348..12366  
 /note="17-42-319.mis complement"  
 primer\_bind 12581..12603  
 /note="17-42..rp complement"  
 exon 12947..12958  
 /number=1  
 exon 13470..13526  
 /number=2  
 exon 13641..13752  
 /number=3  
 exon 14271..15968  
 /number=4  
 primer\_bind 14992..15012  
 /note="17-41..pu"  
 primer\_bind 15222..15240  
 /note="17-41-250.mis"  
 misc\_binding 15229..15253  
 /bound\_moiety="17-41-250.probe"  
 variation 15241  
 /note="17-41-250 : polymorphic base C or T"  
 primer\_bind 15242..15260  
 /note="17-41-250.mis complement"  
 primer\_bind 15460..15482  
 /note="17-41..rp complement"  
 misc\_feature 15969..17969  
 /note="3;regulatory region"  
 primer\_bind 42070..42090  
 /note="20-841..pu"  
 primer\_bind 42199..42217  
 /note="20-841-149.mis"  
 misc\_binding 42206..42230

variation /bound\_moiety="20-841-149.probe"  
 42218  
 /note="20-841-149 : polymorphic base A or C"  
 primer\_bind 42219..42237  
 /note="20-841-149.mis complement"  
 primer\_bind 42572..42591  
 /note="20-841..rp complement"  
 primer\_bind 45328..45347  
 /note="20-842..pu"  
 primer\_bind 45423..45441  
 /note="20-842-115.mis"  
 misc\_binding 45430..45454  
 /bound\_moiety="20-842-115.probe"  
 variation 45442  
 /note="20-842-115 : polymorphic base A or G"  
 primer\_bind 45443..45461  
 /note="20-842-115.mis complement"  
 primer\_bind 45863..45883  
 /note="20-842..rp complement"  
 primer\_bind 76644..76664  
 /note="20-853..pu"  
 primer\_bind 77039..77057  
 /note="20-853-415.mis"  
 misc\_binding 77046..77070  
 /bound\_moiety="20-853-415.probe"  
 variation 77058  
 /note="20-853-415 : polymorphic base C or T"  
 primer\_bind 77059..77077  
 /note="20-853-415.mis complement"  
 primer\_bind 77166..77185  
 /note="20-853..rp complement"  
 BASE COUNT 20291 a 17705 c 19080 g 23919 t 6 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0 Length: 81001  
 Score: 312.00 Matches: 312  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 85.25% Gaps: 0  
 DB: 6  
 US-09-842-364-3 (1-366) x AX469876 (1-81001)  
 QY 55 ThlleuLysAspSerLeuGluGlnAspLeuAsnAsnMetCAsnLysPheLeuGluLysLeu 74  
 |||||  
 Db 14272 ACCCTGAAGACAGCCCTTGAGCAGACCTCAACAAATATGCAACAGTTCCTGCAAAAGCTG 14331  
 |||||  
 QY 75 AcgProLeuSerGlySerGluAlaProArgLeuProGlnAspProValGlyMetArgArg 94  
 |||||  
 Db 14332 AGCCCTCTCAGTGGAGCGGAGCGCTCCGCTCCACAGACCCGCGCATGCGCGG 14391  
 |||||  
 QY 95 GlnLeuGlnGluGluLeuGluGluValLysAlaArgLeuGlnProArgMetAlaGluAla 114  
 |||||  
 Db 14392 CAGCTGCAGAGAGATTGGAGGAGTGTAAGGCTCGCTCCAGGCCCACTGGAGAGGGG 14451  
 |||||  
 QY 115 HisGluLeuValGlyTyrPheLeuGluGluArgGlnGlnLeuLysProTyrThrMet 134  
 |||||  
 Db 14452 CACGAGCTGCTGGCTGGAAATTTGGAGGCTTCCGACACCACTGAAGCCCTTACAGATG 14511  
 |||||  
 QY 135 AspleuMetGluGlnValAlaLeuArgValGlnGluLeuGlnGlnLeuArgValAla 154  
 |||||  
 Db 14512 GATCTGATGAGCAGAGTGGCCCTCGCCCTGCACAGAGCTCAGAGGACGATGGCGGTG 14571  
 |||||  
 QY 155 GlyGluAspThrLysAlaGlnLeuLeuGluGlyValAspGluAlaTrrAlaLeuLeuGln 174  
 |||||  
 Db 14572 GGGGAAGACACCAAGCCAGTGTGCGGGGGGTGAGCAGAGCTTGGCTTGGCGAG 14631  
 |||||  
 QY 175 GlyLeuGlnSerArgValAlaHisHisThrGlyAlaGlyPheGlyGluLeuPheHisProTyr 194  
 |||||  
 Db 14632 GGACTGCAAGCGCGGTGTGACACACAGCCGCTTCAAGAGACTTTCACACCATAC 14691  
 |||||  
 QY 195 AlaGluSerLeuValSerGlyLeuGlyArgHisValGlnGluLeuHisArgSerValAla 214





## COMMENT

and Chemical Research (RIKEN), Genomic Sciences Center (GSC):  
 Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,  
 Japan (E-mail: hatori@gsc.riken.go.jp,  
 URL: http://hgp.gsc.riken.go.jp/, Tel: 81-42-778-9923,  
 Fax: 81-42-778-9924)  
 On May 30, 2000 this sequence version replaced g1:7288175.

----- Genome Center

Center: RIKEN Genomic Sciences Center (GSC)

Center code: RIKEN

Web site: http://hgp.gsc.riken.go.jp/

Contact: hatori@gsc.riken.go.jp

----- Project Information

Center project name: HumDrafl1

Center clone name: CTC-270C21

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads  
 Chemistry: Dye-terminator ET-amersham; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 162298 bases at least Q40  
 Consensus quality: 167630 bases at least Q30  
 Consensus quality: 169842 bases at least Q20  
 Insert size: 172782; sum-of-contigs  
 Quality coverage: 4.96x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 29 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1      16627 contig of 16627 bp in length
16728 30620 contig of 13893 bp in length
30721 42439 contig of 11719 bp in length
42540 53421 contig of 10882 bp in length
53522 64748 contig of 11227 bp in length
64849 76919 contig of 12071 bp in length
77020 85523 contig of 8504 bp in length
85624 94708 contig of 9085 bp in length
94809 104076 contig of 9268 bp in length
104177 112629 contig of 8453 bp in length
112730 119309 contig of 6580 bp in length
119410 126696 contig of 7287 bp in length
126797 132422 contig of 5626 bp in length
132523 137920 contig of 5398 bp in length
138021 142190 contig of 4170 bp in length
142291 145283 contig of 2993 bp in length
145384 148550 contig of 3167 bp in length
148651 151425 contig of 2775 bp in length
151526 155432 contig of 3907 bp in length
155533 158018 contig of 2486 bp in length
158119 160661 contig of 2543 bp in length
160762 163242 contig of 2481 bp in length
163343 165254 contig of 1912 bp in length
165355 167624 contig of 2270 bp in length
167725 169398 contig of 1674 bp in length
169499 171146 contig of 1648 bp in length
171247 173044 contig of 1798 bp in length
173145 174332 contig of 1188 bp in length
174433 175582 contig of 1150 bp in length

```

Sequence updated (26-May-2000).

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 29 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1      16627 contig of 16627 bp in length
16728 30620 contig of 13893 bp in length
30721 42439 contig of 11719 bp in length

```

## FEATURES

source

```

30721 42439: contig of 11719 bp in length
42440 42539: gap of 100 bp
42540 53421: contig of 10882 bp in length
53422 53521: gap of 100 bp
53522 64748: contig of 11227 bp in length
64749 64848: gap of 100 bp
64849 76919: contig of 12071 bp in length
76920 77019: gap of 100 bp
77020 85523: contig of 8504 bp in length
85524 85623: gap of 100 bp
85624 94708: contig of 9085 bp in length
94709 94808: gap of 100 bp
94809 104076: contig of 9268 bp in length
104077 104176: gap of 100 bp
104177 112629: contig of 8453 bp in length
112630 112729: gap of 100 bp
112730 119309: contig of 6580 bp in length
119310 119409: gap of 100 bp
119410 126696: contig of 7287 bp in length
126697 126796: gap of 100 bp
126797 132422: contig of 5626 bp in length
132423 132522: gap of 100 bp
132523 137920: contig of 5398 bp in length
137921 138020: gap of 100 bp
138021 142190: contig of 4170 bp in length
142191 142290: gap of 100 bp
142291 145283: contig of 2993 bp in length
145284 145383: gap of 100 bp
145384 148550: contig of 3167 bp in length
148551 148650: gap of 100 bp
148651 151425: contig of 2775 bp in length
151426 151525: gap of 100 bp
151526 155432: contig of 3907 bp in length
155433 155532: gap of 100 bp
155533 158018: contig of 2486 bp in length
158019 158118: gap of 100 bp
158119 160661: contig of 2543 bp in length
160662 160761: gap of 100 bp
160762 163242: contig of 2481 bp in length
163243 163342: gap of 100 bp
163343 165254: contig of 1912 bp in length
165255 165354: gap of 100 bp
165355 167624: contig of 2270 bp in length
167625 167724: gap of 100 bp
167725 169398: contig of 1674 bp in length
169399 169498: gap of 100 bp
169499 171146: contig of 1648 bp in length
171147 171246: gap of 100 bp
171247 173044: contig of 1798 bp in length
173045 173144: gap of 100 bp
173145 174332: contig of 1188 bp in length
174333 174432: gap of 100 bp
174433 175582: contig of 1150 bp in length.

1. 175582
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="CTC-270C21"
1. 16627
/misc_feature
/note="assembly-fragment"
16728..30620
/misc_feature
/note="assembly-fragment"
30721..42439
/misc_feature
/note="assembly-fragment"
42540..53421
/misc_feature
/note="assembly-fragment"
53522..64748
/misc_feature
/note="assembly-fragment"
64849..76919
/misc_feature
/note="assembly-fragment"
77020..85523

```

```

/misc_feature      /note="assembly_fragment"
                    85624..94708
/misc_feature      /note="assembly_fragment"
                    94809..104076
/misc_feature      /note="assembly_fragment"
                    104177..112629
/misc_feature      /note="assembly_fragment"
                    112730..119309
/misc_feature      /note="assembly_fragment"
                    119410..126696
/misc_feature      /note="assembly_fragment"
                    126797..132422
/misc_feature      /note="assembly_fragment"
                    132523..137920
/misc_feature      /note="assembly_fragment"
                    138021..142190
/misc_feature      /note="assembly_fragment"
                    142291..145283
/misc_feature      /note="assembly_fragment"
                    145384..148550
/misc_feature      /note="assembly_fragment"
                    148651..151425
/misc_feature      /note="assembly_fragment"
                    151526..155432
/misc_feature      /note="assembly_fragment"
                    155533..158018
/misc_feature      /note="assembly_fragment"
                    158119..160661
/misc_feature      /note="assembly_fragment"
                    160762..163242
/misc_feature      /note="assembly_fragment"
                    163343..165254
/misc_feature      /note="assembly_fragment"
                    165355..167624

```

## Alignment Scores:

```

Pred. No.:      0      Length:      175582
Score:          312.00  Matches:      312
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match:      85.25%  Indels: 0
DH:               2      Gaps: 0

```

US-09-842-364-3 (1-366) x AP001481 (1-175582)

```

QY 55 ThrLeuLysASPSeSerLeuCluGlnAspLeuAsnAsnMetAsnLysPheLeuGluLysLeu 74
|||||
Db 106171 ACCCTGAAGAAGACACCCCTTCAGCAGACACCCATATCAACAGTTCCTCGAAAGACCTG
|||||
QY 75 ArgProIeuSerClySerClyValaIProArgIeuProGlnAspProValaIcylMetArgArg 94
|||||
Db 106111 AGGCTCTGAGTGGAGGAGGAGCTCTCGGCTCCACAGGAGCCGCTGGGCTATGGGCGG
|||||
QY 95 GlnLeuGlnGluGlnLeuCluGlnValaLysAlaArgLeuGlnProGlyMetAlaGluAla 114
|||||
Db 106051 CAGCTGCAGAGAGACTTGCAGAGAGCTGAAGCTCCCTCCACCCCTACATGCAGAGCGG
|||||
QY 115 HisGluLeuValaGlyTrpAsnLeuGlnGlyLeuArgGlnGlnLeuLysProGlyThrMet 134
|||||
Db 105991 CACGAGCTGGCTGGCTGGCAATTTGGAGGCGCTTGGCGCAGCAACTGAACCCCTACAGGAGT
|||||
QY 135 AspLeuMetGluGlnValaLysAlaLeuArgValaGlnGluLeuGlnGlnLeuArgVala 154
|||||
Db 105931 GATCTGATGGAGCAGAGTGGCTGGCGTGGAGGAGCTGCAGAGAGCAGTGGCGTGGTGG
|||||
QY 155 GlyIuAspPheTrpLysAlaGlnLeuLeuCluGlyValaLysPheLysAlaTrpAlaLeuLeuGln 174
|||||
Db 105871 GGGGAGAGACACCAAGCCCACTTGGCTGGGCGGCTGCAGAGAGCTTGGCTTCTGCTGCAG
|||||
QY 175 GlyLeuGlnSerArgValaValaHisIsthrGlyArgPheLysGlnLeuPheHisProArg 194
|||||
Db 105811 GGACTGCGAGAGCGCGCTGGTGGTGCACACACGCGCCCTTCAAGAGACTCTTCCACCCATAC
|||||

```

```

QY 195 AlaGluSerLeuValSerClyTlleGlyArgHisValGlnGluLeuHisArgSerVala 214
|||||
Db 105751 GCGGAGAGCTGTGTGAGGAGCTGGGCGCCAGCTGCAGAGACTGCACCGAGTGGCT
|||||
QY 215 ProHisAlaProAlaSerProAlaArgLeuSerArgCysValaGlnValLeuSerArgLys 234
|||||
Db 105691 CCGCAGAGCCCGCCCGCAGCCCGCCCGCCCTCAGCTCCGCTGCAGTGCCTCCCGGAAG
|||||
QY 235 LeuThrLeuLysAlaLysAlaLeuHisAlaArgIleGlnGlnAsnLeuAspGlnLeuArg 254
|||||
Db 105631 CTCAGCGCTCAAGGCAAGGCGCTGCAGCGCAGCATTCAGCAAGCTGCAGCAGCTGGCG
|||||
QY 255 GluGluLeuSerArgAlaPheAlaGlyTrpGlyThrGluGluGluValaGlyProAspPro 274
|||||
Db 105571 GAAAGAGCTAGCAGAGAGCTTTCAGAGCAGTGGAGTGAAGAGGGCGCGCCGAGACCC
|||||
QY 275 GlnMetLeuSerGluGluValaArgGlnArgLeuGlnAlaPheArgGlnAspThrTyLeu 294
|||||
Db 105511 CAGAGTCTCTCCGAGAGAGCTGGCCAGCCAGCTTCAGAGCTTCCCGCAGAGCCTAACCTG
|||||
QY 295 GlnIleAlaAlaPheThrArgAlaIleAspGlnGluThrGluGluValaGlnGlnLeu 314
|||||
Db 105451 CAGATAGCTGCTTACTGCTGCGCCATTCAGCAGAGAGTGAAGAGAGTGCAGAGCAGCTG
|||||
QY 315 AlaProProProProGlyHisSerAlaPheAlaProGluPheGlnGlnThrAspSerGly 334
|||||
Db 105391 GCGCCACCTCCACAGGAGCAGAGCTGCTGCGCCAGAGTTCACAAACAGCAGTGGC
|||||
QY 335 LysValLeuSerLysLeuGlnAlaArgLeuAspAspLeuTrpGlnAspLysIleThrIstSer 354
|||||
Db 105331 AAGCTTCTCAGCAGACTGCAGGCGCCGCTGCGAGTCTGCGAGAGATCATCTCAGACG
|||||
QY 355 LeuHisAspGlnGlyHisSerHisLeuGlnAspPro 366
|||||
Db 105271 CTTCAATGACGAGGCGCAGACGATCTGGGGGAGACCC 105236

```

```

RESULT 15
AC007707/c 188971 bp DNA linear PRI 25-NOV-1999
LOCUS      Homo sapiens chromosome 11 clone 442e11 from RPC11 library map
DEFINITION 11q23, complete sequence.
ACCESSION  AC007707
VERSION     AC007707.13 GI:5764724
KEYWORDS   HTC.
SOURCE      Homo sapiens.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE  1 (bases 1 to 188971)
            Shaikh,T.H., Budarf,M.L., Celic,L., Zackai,E.H. and Emanuel,B.S.
            Clustered 11q23 and 22q11 breakpoints ad 3:1 meiotic malesegregations
            in multiple unrelated t(11;22) families
            Am. J. Hum. Genet. (1999) In press
            2 (bases 1 to 188971)
            Hu,P. and Roe,B.A.
            Direct Submission
            Submitted (02-JUN-1999) Department of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
            3 (bases 1 to 188971)
            Hu,P. and Roe,B.A.
            Direct Submission
            Submitted (25-AUG-1999) Department of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
            4 (bases 1 to 188971)
            Hu,P. and Roe,B.A.
            Direct Submission
            Submitted (26-AUG-1999) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
            5 (bases 1 to 188971)
            Hu,P. and Roe,B.A.

```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 23, 2003, 17:46:52 : Search time 2033 Seconds

(without alignments)  
2915.666 Million cell updates/sec

Title: US-09-842-364-3

Perfect score: 366  
Sequence: 1 MASMAVLTWALALISA...LMEDITHSLHDOCHSLGDP 366

Scoring table:

OLIGO  
Xgapop 60.0, Xgapext 60.0  
Ygapop 60.0, Ygapext 60.0  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Word size: 6

Total number of hits satisfying chosen parameters: 231638

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Command line parameters:  
-MODEL-frame+ p2n.model -DEV-xml  
-O/cgn2.1/USPTO.spool/US09842364/runat.16012003.164829.10661/app.query.fasta.1.519  
-DB-PST -QFMT-fastap -SUFFIX-olig.rst -MINMATCH=0.1 -IDOPCI=0 -IDOPERT=0  
-UNITS-bits -START=1 -END=1 -MATRIX-oligo -TRANS-human40.cdi -LIST=45  
-DOCLION=200 -THR\_SCORE=quality -THR\_MIN=6 -ALION=15 -MODE-LOCAL -OUTFMT-pto  
-NORR-ext HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09842364 -CGN.1.1349 -runat.16012003.164829.10661 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAL -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7  
-YGAOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysts of the total score distribution.

#### SUMMARIES

Result	No.	Score	Match	Length	DB	ID	Description
c	1	75	20.5	696	17	AC097513	AC097513 Pan trogl
c	2	26	7.1	524	9	A1265480	A1265480 u 02e07.x
c	3	26	7.1	542	12	Bf659944	Bf659944 ma25e08.
c	4	26	7.1	549	13	B1148686	B1148686 602912117
c	5	26	7.1	554	9	A1528320	A1528320 u19e404.y
c	6	26	7.1	606	13	B1144716	B1144716 602909506
c	7	26	7.1	608	12	Bf384518	Bf384518 602046482
c	8	26	7.1	612	13	B1221610	B1221610 602936921
c	9	26	7.1	612	9	AA987093	AA987093 uc74f12.y
c	10	26	7.1	617	12	Bf384624	Bf384624 602046895
c	11	26	7.1	645	12	Bf234067	Bf234067 602024986
c	12	26	7.1	648	12	Bf385873	Bf385873 602046040
c	13	26	7.1	660	12	Bf385843	Bf385843 602046010
c	14	26	7.1	663	13	B1220419	B1220419 602935137
c	15	26	7.1	672	12	Bf384592	Bf384592 602046861
c	16	26	7.1	672	13	B1332079	B1332079 602981411
c	17	26	7.1	699	12	Bf385663	Bf385663 602047318
c	18	26	7.1	705	13	B1144468	B1144468 602908638
c	19	26	7.1	705	13	B1255671	B1255671 602979053
c	20	26	7.1	716	13	B1145333	B1145333 602910476
c	21	26	7.1	722	12	Bf384847	Bf384847 602046343
c	22	26	7.1	722	12	Bf385520	Bf385520 602047065
c	23	26	7.1	727	12	Bf383715	Bf383715 602044622
c	24	26	7.1	735	13	B1145765	B1145765 602909228
c	25	26	7.1	746	13	B1331565	B1331565 602908983
c	26	26	7.1	758	12	Bf237009	Bf237009 602026641
c	27	26	7.1	762	12	Bf233452	Bf233452 602024241
c	28	26	7.1	767	13	B1149103	B1149103 602911511
c	29	26	7.1	767	13	B1246578	B1246578 602958496
c	30	26	7.1	769	13	B1144954	B1144954 602909606
c	31	26	7.1	773	13	B1145650	B1145650 602910686
c	32	26	7.1	778	13	B1145877	B1145877 602911175
c	33	26	7.1	789	13	B1247725	B1247725 602959363
c	34	26	7.1	791	12	Bf233098	Bf233098 602023839
c	35	26	7.1	791	13	B1147941	B1147941 602912655
c	36	26	7.1	792	13	B1145009	B1145009 602909669
c	37	26	7.1	792	13	B1146191	B1146191 602911005
c	38	26	7.1	793	13	B1144673	B1144673 602909969
c	39	26	7.1	798	12	Bf234162	Bf234162 602026017
c	40	26	7.1	804	13	B1145831	B1145831 602911107
c	41	26	7.1	805	13	B1148123	B1148123 602912487
c	42	26	7.1	806	13	B1148302	B1148302 602914316
c	43	26	7.1	809	12	Bf233225	Bf233225 602023779
c	44	26	7.1	815	13	B1331154	B1331154 602984189
c	45	26	7.1	820	13	B1332727	B1332727 602984481

#### ALIGNMENTS

RESULT 1  
AC097513/c  
LOCUS AC097513  
DEFINITION Pan troglodytes DNA, clone: PTB-099D09.F, genomic survey sequence.  
ACCESSION AC097513  
VERSION AC097513.1 GI:16718030  
KEYWORDS GSS.  
SOURCE Pan troglodytes male lymphoblast DNA, clone: PTB-099D09.F, genomic survey sequence.  
ORGANISM Pan troglodytes  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.  
REFERENCE  
1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.



TITLE	BAC end sequences of library P7B
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 696)
AUTHORS	Fujiyama, A., Hallori, M., Toyoda, A., Taylor, T.D., Yada, T., Tokoi, Y., Watanabe, H. and Sakaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimpbes@gsc.riken.go.jp, URL: <a href="http://ngp.gsc.riken.go.jp/">http://ngp.gsc.riken.go.jp/</a> , Tel: 81-45-503-9111, Fax: 81-45-503-9170)
COMMENT	Clones are derived from the chimpanzee BAC library P7B This BAC end was generated during the Rad process and may have higher chance of clone tracking errors.
PRIMERS	

```

Sequencing: -21M13
LIBRARY
  Vector      : pKS145
  R.Site 1    : SacI
  R.Site 2    : SacI
  Location/Qualifiers
    1..696

```

```

/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-099D09.F"
/sex="male"
/cell_type="lymphoblast"
/clone_1fb="PTB Chimpanzee Male BAC Library"
BASE COUNT
128 a 250 c 180 g 138 t
ORIGIN

```

Alignment Scores:	
Pred. No.:	4,13e-59
Score:	75.00
Percent Similarity:	100.00%
Best local Similarity:	100.00%
Query Match:	20.49%
DB:	17
Length:	636
Matches:	75
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-09-842-364-3 (1-366) x AG097513 (1-696)

QY	95	GlnLeuGngInGluLeuGluGluValLysAlaArgLeuGInProTyrMetAlaGluAla	114
Db	368	CAGCTGCAGAGAGATTGAGAGAGGTGAAGCGCTCCGCTCCAGCCCTACATGGCAGAGCGC	300
QY	115	HisGluLeuValGlyTyrPasnLeuGluGlyLeuArgGngInLeuLysProTyrThrMet	134
Db	308	CACGAGCTGGCTGGCTCGAAATTTGCAGCGCTTCGCGCAGCAACCTACACGATCG	249
QY	135	AspLeuMetGluGlnValAlaIleuArgValGlnGlnLeuGngInGluGlnLeuArgValVal	154
Db	248	GATCTGATGAGCAGAGGTGGCCCTCGGCTGGAGGAGCTGCAGGAGCAGATTGGCGGTGGTG	189

QY 155 GlytluAspPrhrllysAlaInleuLeuGlyGlyValAspClnuAla 165  
 |||||  
 Db 188 GGAGAAGACACCAAGCCCAAGCTGCTGGGGGGCTGGACACAGGGC 194

RESULT 2	LOCUS	DEFINITION	ACCESSION
AI265480/c	AI265480	524 bp	mus mus
	U102607.x1	Sugano mouse liver mRNA	mus mus
	IMAGE:1890756	3', mRNA sequence.	mus mus

ACCESSION	AI265480
VERSION	AI265480.1
KEYWORDS	GI:3883638
SOURCE	EST.
ORGANISM	house mouse.
	house mouse.
	house mouse.

ORGANISM  
MUS MUSCULUS  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 524)  
Marrá, M., Hillier, T., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,

rhelsing, B., Wyle, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
TITLE The MashU-HIMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT Contact: Maria M/Mouse EST Project

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel.: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LINL: core@imac-consortium.info@imgc.linl.gov for further info.  
MGI:975080  
mouseest powered image, not found

Seq primer: custom primer used  
High quality sequence stop: 487.  
Location/Qualifiers

FEATURES	Location/Qualifiers
source	1. .524

```

/organism="Mus musculus"
/slibrary="C57BL"
/db_xref="taxon:10090"
/clone_image="IMAGE:1890756"
/clone_lib="Sugano mouse liver mlia"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/notes="Organ: liver. Vector: pME18s-Flu3. Site.1: DraIII
(CACTGTGTG); Site.2: DraIII (CAGCATGTG); 1st strand cDNA
was primed with an oligo(dT) primer
[ATGTCGCGCTTTTCTTTTCTTTTCTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor (TCTTCGCGCTACTGC), digested
and cloned into distinct DraIII sites of the pME18s-Flu3
vector (5' site CACTGTGTG, 3' site CAGCATGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTTCGCTCTCTAAACCTCGC and 3' end
primer CACCTGCAGCTCGAGCA."

```

BASE COUNT	113 a	160 c	131 g	120 l
ORIGIN				
Alignment Scores:				
Pred. No.:	6.16e-14	Length:		
Score:	26.00	Matches:		
Percent Similarity:	100.00%	Conservative:		
Best local Similarity:	100.00%	Mismatches:		
Query Match:	7.10%	InucIs:		
DB:	9	Caps:		
US-09-842-364-3 (1-366) x AI265480 (1-524)				

QY 144 ValGInGInLeuGInGInLeuArgValValGInGInAspThrLysAlaGInLeuLeu 1637  
|||||  
Db 517 GTTCAGGAGCTGCAGGAGCACACTCGGTGTGGAGACAGACCAAGCGTCACCTCTG 458B

QY	164	GLYGLYValAspGluAla	169
nb	457	GGGGCGGTGGACGAGCGG	440

RESULT 3	LOCUS	DEFINITION
BF659944	542 bp	mus musculus cDNA clone IMAGE:3812078 5'
BF659944		similar to TR:090UH3 090UH3 REGENERATION ASSOCIATED PROTEIN 3. 1

ACCESSION	mRNA sequence.
VERSION	BF659944
KEYWORDS	BF659944.1 GI:11925078
SOURCE	EST.
ORGANISM	house mouse.
	Mus musculus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;





## Alignment Scores:

Pred. No.: 7.05e-14 Length: 608  
 Score: 26.00 Matches: 26  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 7.10% Indels: 0  
 DB: 12 Caps: 0

US-09-842-364-3 (1-366) x BF384518 (1-608)

OY 144 ValGInGluLeuGInGluGInLeuArgValValGlyGlnAspThrLysAlaGlnLeu 163  
 |||||||  
 DB 424 CTGCAGAGCTGCAGACACACTCCCTGCTGCAGACAGACACCAAGCTCAGCTCTG 483  
 |||||||  
 OY 164 GYGlyValAspGlnAla 169  
 |||||||  
 DB 484 GGGGGCTGGAGCAGCGC 501

## RESULTS

BI221610

LOCUS 608 bp mRNA linear EST 11-JUL-2001

DEFINITION 602936921F1 NCL\_CGAP\_Li9 Mus musculus cDNA clone IMAGE:5100364 5',

ACCSSION BI221610

VERSION BI221610.1 GI:14675054

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

TITLE NIH-MGC http://mgs.nci.nih.gov/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgaps-f@mail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.lnl.gov

Plate: LLM11241 row: 1 column: 05

High quality sequence stop: 608.

Location/Qualifiers

1..608

/organism="Mus musculus"

/strain="FVH/N"

/db\_xref="taxon:10090"

/clone="IMAGE:5100364"

/clone\_1lb="NCI CGAP\_Li9"

/lab\_host="DH10B (TI phage-resistant)"

/note="Organ: Liver; Vector: PCMV-SPOrt6; Site\_1: NotI;

Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.9 kb. Constructed by Life

Technologies. Note: this is a NCI CGAP Library."

BASE COUNT 149 a 162 c 203 g 94 t

ORIGIN

## Alignment Scores:

Pred. No.: 7.05e-14 Length: 608  
 Score: 26.00 Matches: 26  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 7.10% Indels: 0  
 DB: 13 Caps: 0

US-09-842-364-3 (1-366) x BI221610 (1-608)

OY 144 ValGInGluLeuGInGluGInLeuArgValValGlyGlnAspThrLysAlaGlnLeu 163  
 |||||||  
 DB 435 CTGCAGAGCTGCAGACACACTCCCTGCTGCAGACAGACACCAAGCTCAGCTCTG 494

OY 164 GYGlyValAspGlnAla 169  
 |||||||  
 DB 495 GGGGGCTGGAGCAGCGC 512

## RESULTS

AA987093 612 bp mRNA linear EST 28-MAY-1998

LOCUS uc74112.y1 Sugano mouse liver mla Mus musculus cDNA clone

DEFINITION IMAGE:1431407 5' similar to SW:AP4\_MACPA P33621 APOLOPROTEIN

A-IV PRECURSOR ; mRNA sequence.

ACCSSION AA987093

VERSION AA987093.1 GI:3167856

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

TITLE Matra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,

Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,

Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and

Waterston, R.

JOURNAL The Washu-HMI Mouse EST Project

COMMENT Unpublished (1996)

Contact: Matra M/Mouse EST Project

Washu-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

MGI:915475

Seq primer: primer name ambiguous

High quality sequence stop: 508.

Location/Qualifiers

1..612

/organism="Mus musculus"

/strain="C57BL"

/db\_xref="taxon:10090"

/clone="IMAGE:1431407"

/clone\_1lb="Sugano mouse liver mla"

/sex="female"

/dev\_stage="adult"

/lab\_host="DH10B"

/note="Organ: liver; Vector: pME18S-Fl3; Site\_1: DraIII

(CACCTGTCG); Site\_2: DraIII (CACCTGTCG); 1st strand cDNA

was primed with an oligo(dT) primer

(ATGTGGCTTTTCTTTTCTTTT); double-stranded cDNA was

ligated to a DraIII adaptor (TGTGGCTGCTG), digested

and cloned into distinct DraIII sites of the pME18S-Fl3

vector (5' site CACCTGTCG, 3' site CACCTGTCG). XhoI should

be used to isolate the cDNA insert. Size selection was

performed to exclude fragments <1.5kb. Library

constructed by Dr. Sumio Sugano (University of Tokyo

Institute of Medical Science). Custom primers for

sequencing: 5' end primer CTTCGCTCTTAAAGCTGCG and 3' end

primer CGACCTGACCTGACGACAC.

BASE COUNT 152 a 166 c 202 g 92 t

## ORIGIN

## Alignment Scores:

Pred. No.: 7.1e-14 Length: 612  
 Score: 26.00 Matches: 26  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 7.10% Indels: 0  
 DB: 9 Caps: 0

US-09-842-364-3 (1-366) x AA987093 (1-612)

```

QY 144 ValGIngluLeuGIngluGInleuAryValGlyIuAspThrIlySAlaGInleu 163
DB 443 GTGCAGGAGCTGCAGACACCTGCGCTGCTGGAGAAACACCAAGCTCAGCTCTG 502
QY 164 GlyIyValAspGluA1a 169
DB 503 GGGGGCGTGCAGAGCGC 520

RESULT 10
BF384624 617 bp mRNA linear EST 27-NOV-2000
LOCUS 602046895F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:4196542 5'
DEFINITION mRNA sequence.
ACCESSION BF384624.1 GI:11365929
VERSION BF384624
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 617)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM9533 row: m column: 23
High quality sequence stop: 615.
Location/Qualifiers
1. 617
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image="4196542"
/clone_id="NCI_CGAP_L19"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: liver; Vector: pCMV-Sport6; Site: 1; Note:
Site: 2; Salt: Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 152 a 164 c 206 g 95 t
ORIGIN
1
Alignment Scores:
Pred. No.: 7,15e-14 Length: 617
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.10% Indels: 0
DB: 12 Gaps: 0
US-09-842-364-3 (1-366) x BF384624 (1-617)
QY 144 ValGIngluLeuGIngluGInleuAryValGlyIuAspThrIlySAlaGInleu 163
DB 423 GTGCAGGAGCTGCAGACACCTGCGCTGCTGGAGAAACACCAAGCTCAGCTCTG 482
QY 164 GlyIyValAspGluA1a 169
DB 483 GGGGGCGTGCAGAGCGC 500

RESULT 11
BF234067 645 bp mRNA linear EST 14-NOV-2000
LOCUS 602024986F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:4160397 5'
DEFINITION

```

```

ACCESSION BF234067
VERSION BF234067.1 GI:11145152
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 645)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM9439 row: k column: 22
High quality sequence stop: 639.
Location/Qualifiers
1. 645
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image="4160397"
/clone_id="NCI_CGAP_L19"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: liver; Vector: pCMV-Sport6; Site: 1; Note:
Site: 2; Salt: Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 153 a 180 c 212 g 100 t
ORIGIN
1
Alignment Scores:
Pred. No.: 7,44e-14 Length: 645
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.10% Indels: 0
DB: 12 Gaps: 0
US-09-842-364-3 (1-366) x BF234067 (1-645)
QY 144 ValGIngluLeuGIngluGInleuAryValGlyIuAspThrIlySAlaGInleu 163
DB 413 GTGCAGGAGCTGCAGACACCTGCGCTGCTGGAGAAACACCAAGCTCAGCTCTG 472
QY 164 GlyIyValAspGluA1a 169
DB 473 GGGGGCGTGCAGAGCGC 490

RESULT 12
BF385873 648 bp mRNA linear EST 27-NOV-2000
LOCUS 602046040F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:4198808 5'
DEFINITION mRNA sequence.
ACCESSION BF385873
VERSION BF385873.1 GI:11367178
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 648)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

```

Email: c9apbs-remail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1AM9531 row: 0 column: 09  
High quality sequence stop: 646.  
Location/Qualifiers

/strain="FVB/N"  
/db.xref="taxon:10090"  
/clone="IMAGE:4195514"  
/clone\_lib="NCI\_CGAP\_L19"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: Liver; Vector: pCMV-SPORT6; Site: 1; NotI; Site-2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.9 kb. Constructed by Life Technologies. Note: this is a NCI\_CGAP library."  
BASE COUNT 157 a 185 c 217 g 101 t  
ORIGIN

Alignment Scores:  
Pred. No.: 7.48e-14 Length: 648  
Score: 26.00 Matches: 26  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 7.10% Indels: 0  
DB: 12 Gaps: 0  
US-09-842-364-3 (1-366) x BF385843 (1-648)  
LOCUS 660 bp mRNA linear EST 27-NOV-2000  
DEFINITION 602046010P1 NCI\_CGAP\_L19 Mus musculus cDNA clone IMAGE:4195514 5',  
mRNA sequence.  
ACCESSION BF385843  
VERSION BF385843.1 GI:11367148  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
NIH-MGC http://mgi.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: c9apbs-remail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1AM9531 row: c column: 03  
High quality sequence stop: 658.  
Location/Qualifiers

/strain="FVB/N"  
/db.xref="taxon:10090"  
/clone="IMAGE:4195514"  
/clone\_lib="NCI\_CGAP\_L19"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: Liver; Vector: pCMV-SPORT6; Site: 1; NotI; Site-2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.9 kb. Constructed by Life Technologies. Note: this is a NCI\_CGAP library."  
BASE COUNT 157 a 185 c 217 g 101 t  
ORIGIN

Alignment Scores:  
Pred. No.: 7.48e-14 Length: 648  
Score: 26.00 Matches: 26  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 7.10% Indels: 0  
DB: 12 Gaps: 0  
US-09-842-364-3 (1-366) x BF385843 (1-648)  
LOCUS 660 bp mRNA linear EST 27-NOV-2000  
DEFINITION 602046010P1 NCI\_CGAP\_L19 Mus musculus cDNA clone IMAGE:4195514 5',  
mRNA sequence.  
ACCESSION BF385843  
VERSION BF385843.1 GI:11367148  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
NIH-MGC http://mgi.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: c9apbs-remail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1AM9531 row: c column: 03  
High quality sequence stop: 658.  
Location/Qualifiers

/strain="FVB/N"  
/db.xref="taxon:10090"  
/clone="IMAGE:4195514"  
/clone\_lib="NCI\_CGAP\_L19"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: Liver; Vector: pCMV-SPORT6; Site: 1; NotI; Site-2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.9 kb. Constructed by Life Technologies. Note: this is a NCI\_CGAP library."  
BASE COUNT 163 a 183 c 216 g 101 t  
ORIGIN

Alignment Scores:  
Pred. No.: 7.48e-14 Length: 648  
Score: 26.00 Matches: 26  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 7.10% Indels: 0  
DB: 12 Gaps: 0  
US-09-842-364-3 (1-366) x BF385843 (1-648)  
LOCUS 660 bp mRNA linear EST 27-NOV-2000  
DEFINITION 602046010P1 NCI\_CGAP\_L19 Mus musculus cDNA clone IMAGE:4195514 5',  
mRNA sequence.  
ACCESSION BF385843  
VERSION BF385843.1 GI:11367148  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
NIH-MGC http://mgi.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: c9apbs-remail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1AM9531 row: c column: 03  
High quality sequence stop: 658.  
Location/Qualifiers

/strain="FVB/N"  
/db.xref="taxon:10090"  
/clone="IMAGE:4195514"  
/clone\_lib="NCI\_CGAP\_L19"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: Liver; Vector: pCMV-SPORT6; Site: 1; NotI; Site-2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.9 kb. Constructed by Life Technologies. Note: this is a NCI\_CGAP library."  
BASE COUNT 163 a 183 c 216 g 101 t  
ORIGIN

Alignment Scores:  
Pred. No.: 7.48e-14 Length: 648  
Score: 26.00 Matches: 26  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 7.10% Indels: 0  
DB: 12 Gaps: 0  
US-09-842-364-3 (1-366) x BF385843 (1-648)  
LOCUS 660 bp mRNA linear EST 27-NOV-2000  
DEFINITION 602046010P1 NCI\_CGAP\_L19 Mus musculus cDNA clone IMAGE:4195514 5',  
mRNA sequence.  
ACCESSION BF385843  
VERSION BF385843.1 GI:11367148  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
NIH-MGC http://mgi.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: c9apbs-remail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1AM9531 row: c column: 03  
High quality sequence stop: 658.  
Location/Qualifiers

/strain="FVB/N"  
/db.xref="taxon:10090"  
/clone="IMAGE:4195514"  
/clone\_lib="NCI\_CGAP\_L19"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: Liver; Vector: pCMV-SPORT6; Site: 1; NotI; Site-2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.9 kb. Constructed by Life Technologies. Note: this is a NCI\_CGAP library."  
BASE COUNT 163 a 183 c 216 g 101 t  
ORIGIN

Alignment Scores:  
Pred. No.: 7.48e-14 Length: 648  
Score: 26.00 Matches: 26  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 7.10% Indels: 0  
DB: 12 Gaps: 0  
US-09-842-364-3 (1-366) x BF385843 (1-648)  
LOCUS 660 bp mRNA linear EST 27-NOV-2000  
DEFINITION 602046010P1 NCI\_CGAP\_L19 Mus musculus cDNA clone IMAGE:4195514 5',  
mRNA sequence.  
ACCESSION BF385843  
VERSION BF385843.1 GI:11367148  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
NIH-MGC http://mgi.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: c9apbs-remail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1AM9531 row: c column: 03  
High quality sequence stop: 658.  
Location/Qualifiers

/strain="FVB/N"  
/db.xref="taxon:10090"  
/clone="IMAGE:4195514"  
/clone\_lib="NCI\_CGAP\_L19"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: Liver; Vector: pCMV-SPORT6; Site: 1; NotI; Site-2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.9 kb. Constructed by Life Technologies. Note: this is a NCI\_CGAP library."  
BASE COUNT 163 a 183 c 216 g 101 t  
ORIGIN

FEATURES

Source

Alignment Scores:





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 23, 2003, 17:46:02 ; Search time 262 Seconds

(without alignments)  
3145.921 Million cell updates/sec

Title: US-09-842-364-3

Perfect score: 366  
Sequence: 1 MASMAVLVWALALJSAFSA.....LMEDITHSLDQSHLGD 366

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delcext 7.0

Searched: 2185239 seqs, 112599159 residues

Word size: 6

Total number of hits satisfying chosen parameters: 29428

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODEL-frame+ p2n model -DEV-xlh  
-O-Cgq2\_L/USPRO.spool/US09842364/runat.16012003.164829.10640/app\_query.fasta.1.519  
-DB-N\_Geneseq\_101002 -QFMT-fastap -SUFFIX-olig.rng -MIMATCH-0.1 -LOOPL-0  
-LOOPEXT-0 -UNITS-bits -START-1 -END-1 -MATRIX-oligo -TRANS-human40.cdi  
-LIST-45 -DOCALLIGN-200 -THR.SCORER-quality -THR\_MIN-6 -ALIGN-15 -MODE-LOCAL  
-OUTFMT-pto -NORM-ext -HEAPSIZE-500 -MINLEN-0 -MAXLEN-200000000  
-USRR-US09842364.@CGN.1.187.@runat.16012003.164829.10640 -NCPU-6 -ICPU-3  
-NO\_XLPXY -NO\_MMAP -LARGOUDERY -NEG\_SCORES-0 -WAIT -LONGLOG -DEV\_TIMEOUT-120  
-WARN\_TIMEOUT-30 -THREADS-1 -XGAPOP-60 -XGAPEXT-60 -FCGAPOP-6 -FCGAPEXT-7  
-YCAPOP-60 -YCAPEXT-60 -DELOP-6 -DELEXT-7

Database :  
1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT:\*  
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*  
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	366	100.0	1322	21 AAZ45874	Nucleotide sequenc
2	366	100.0	1858	22 AAD19216	Human CG122 (or C8
3	366	100.0	1859	22 AAS22859	Human cDNA encodin
4	366	100.0	1867	21 AAZ45875	Nucleotide sequenc
5	366	100.0	1877	22 AAD19236	Human lipid metabo
6	366	100.0	1877	22 AAS22623	Human cDNA encodin
7	366	100.0	1879	22 AAF30037	Human apolipoprote
8	366	100.0	1894	21 AAB87740	Human secreted pro
9	366	100.0	1894	22 AAF84022	cDNA encoding huma
10	312	85.2	2249	22 AAK89395	Human digestive sy
11	312	85.2	2250	22 AAK89396	Human digestive sy
12	312	85.2	2250	22 AAK89397	Human digestive sy
13	312	85.2	5381	22 AAF30037	Apolipoprotein A-1
14	312	85.2	81001	22 AAF30035	Human apolipoprote
15	279	76.2	1371	24 ABA40279	cDNA encoding huma
16	279	76.2	1371	24 ABA40279	cDNA encoding huma
17	279	76.2	1393	22 AAF97913	Human secreted pro
18	213	58.2	1334	22 AAC85065	Atherosclerosis-as
19	176	48.1	592	22 AAF80605	Receptor #93 parti
20	70	19.1	468	22 AAF97960	Human secreted pro
21	68	18.6	2049	23 AAS73514	Human encoding novel
22	33	9.0	331	22 AAF97961	Human secreted pro
23	33	9.0	558	22 AAK87918	Human digestive sy
24	16	4.4	1282	21 AAZ45872	Nucleotide sequenc
25	16	4.4	1834	21 AAZ45873	Nucleotide sequenc
26	16	4.4	1834	21 AAZ45873	Rat RAP3 gene. Ka
27	14	3.8	831	23 AAS73513	DNA encoding novel
28	11	3.0	5377	24 ABA28379	cDNA transcription
29	9	2.5	1376	21 AAC38218	Arabidopsis thalia
30	9	2.5	1743	24 AAS20017	Partial cDNA encod
31	9	2.5	2850	23 AAS3767	Helicobacter pylor
32	8	2.2	5377	24 ABA28380	Unit of repeated s
33	8	2.2	48	15 AAO57654	Human spliced tran
34	8	2.2	60	24 ABA28379	Human dopamine D4
35	8	2.2	105	16 AAT40295	Non-overlapping re
36	8	2.2	105	20 AAV71234	Human dopamine D4
37	8	2.2	201	16 AAT40296	Non-overlapping re
38	8	2.2	201	20 AAV71235	Non-overlapping re
39	8	2.2	279	22 AAS03067	Human diagnostic a
40	8	2.2	286	21 AAB87100	Rat hepatocytic car
41	8	2.2	345	16 AAT40297	Human dopamine D4
42	8	2.2	345	20 AAV71236	Non-overlapping re
43	8	2.2	346	21 AAC03225	Human secreted pro
44	8	2.2	397	19 AAV60031	Nucleic acid G127
45	8	2.2	431	23 AAS68995	DNA encoding novel

## ALIGNMENTS

RESULT 1  
AAZ45874 standard: cDNA: 1322 BP.

AAZ45874:  
25-APR-2000 (first entry)  
Nucleotide sequence of the 1322 bp cDNA sequence of human RAP3 gene.  
RAP3; regeneration association protein 3; liver regeneration;  
liver proliferation; ss.  
Homo sapiens.

Key Location/Qualifiers  
19..1110  
/\*tag- a



```

FT      /tag= a          /*product= "Human CG122 (or C868) protein"
FT      sig_peptide    46..114
FT      mat_peptide     /tag= b
FT      115..1143
FT      /*tag= c
FT      /*product= "Human mature CG122 (or C868) protein"
XX
XX      WO200179446-A2.
XX
XX      PD
XX      25-OCT-2001.
XX
XX      PF
XX      16-APR-2001; 2001WO-USJ2529.
XX
XX      PR
XX      14-APR-2000; 2000US-197137P.
PR      20-JUN-2000; 2000US-0598042.
PR      03-AUG-2000; 2000US-0631451.
PR      22-SEP-2000; 2000US-0667298.
PR      17-NOV-2000; 2000US-0714936.
XX
XX      (HYSE-) HYSEQ INC.
XX
XX      PA
XX      Ballinger DG, Loeb D, Montgomery JR, Tang TY, Zhou P, Goodrich R;
PI      Liu C, Asundi V, Zhao QA, Wehman T, Drmanac RT, Ken F, Qian XB;
PI      Wang D;
DR      WPI: 2001-611724/70.
DR      P-PSDB: AAEL1922.
XX
XX      Nucleic acids encoding human apolipoproteins, lipases, and lipoprotein'
PT      receptor polypeptides, useful for preventing diagnosing and treating
PT      lipid metabolism disorders, thrombosis and cardiovascular diseases -
XX
XX      Claim 1; Page 146-148; 266pp; English.
XX
XX      The invention relates to polynucleotides encoding proteins CG122, CG179
CC      CG95, CG121, CG162, CG37 CG153 and CG168 which are related to proteins
CC      involved in lipid metabolism and cardiovascular disease such as human
CC      apolipoproteins, lipases and lipoprotein receptor proteins. These DNA
CC      and protein sequences are useful for treating or preventing disorders
CC      associated with apolipoproteins, lipases and lipoprotein receptor (ALRP
CC      expression and for treating lipid metabolism, cardiovascular diseases
CC      and thrombosis. Antibodies against these proteins are useful for
CC      determining the presence of or predisposition to a disease associated
CC      with altered levels of these sequences. ALRP polypeptides are also
CC      useful for identifying agents (agonists and antagonists) that bind to
CC      them and cells expressing ALRP proteins are useful for identifying a
CC      therapeutic agent for use in treatment of a pathology related to
CC      aberrant expression or physiological interactions of this polypeptide.
CC      Vectors comprising these DNA and protein sequences are also useful for
CC      producing ALRP proteins. The nucleic acids and polypeptides of the
CC      invention are also useful for the treatment of occlusive cardiovascular
CC      diseases, myocardial infarction, cerebral ischaemia, angina, arterial
CC      thrombosis, coronary artery thrombosis and cerebral artery thrombosis
CC      or intracardiac thrombosis and stroke. The nucleotides of the invention
CC      are used in gene therapy. The present sequence is human CG122 (or C868)
XX      DNA.
XX
SQ      Sequence 1858 BP; 379 A; 549 C; 582 G; 347 T; 1 other:
XX
Alignment Scores:
Pred. No.: 0 Length: 1858
Score: 366.00 Matches: 366
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatched: 0
Query Match: 100.00% Indels: 0
Db: 22 Gaps: 0
US-09-842-364-3 (1-366) x AAD19216 (1-1858)
OY      1 MetlaasrMetAlaIalValleuThrTPalaLeuAlaLeuSeRaAlapheSeRaLa 20
DB      46 ATGGCAACATAGGCGCGTCGTACCTACCGGCGTCTTGCTCTTCACGCATTTCGGGCC 105

```

RESULT 3	
AA522859	
ID	AA522859 standard; cDNA; 1859 BP.
XX	
AC	AA522859;
XX	
DT	24-OCT-2001 (first entry)
XX	
DE	Human cDNA encoding a novel human protein #425.
XX	
KW	Human; novel protein; ss; Antinaemic; osteopathic; antiinflammatory;
KW	immunomodulatory; cytoskeletal; neuroprotective; vulnery; nootropic;
KW	anticonvulsant; antiallergic; cerebroprotective; antifungal; antiviral;
KW	antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
KW	thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
KW	Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
XX	tissue regeneration; immune disorder.





DB 919 TTGGACAGCAGTGGAGTGAAGGGGGCGGCGGACCCAGATGCTCTCGAGAG 978  
 QY 281 ValArgGlnArgLeuGlnAlaPheArgGlnAspThrTyrLeuGlnIleAlaAlaPheThr 300  
 DB 979 GTGGCGCAGGAGCTTCAGGCTTCCGCGCAGACACCTGACATGAGTGGCTTCACT 1038  
 QY 301 ArgAlaIleAspGlnIleThrGluGluValGlnGlnIleuAlaPheProPheProGly 320  
 DB 1039 CGCGCCATCGACACGACAGTGCAGAGCTGCAGAGCTGGCGCCACCTCCACCGAGCC 1098  
 QY 321 HisSerAlaPheAlaProGluPheGlnGlnThrAspSerGlyValLeuSerLysLeu 340  
 DB 1099 CACAGTGCCTTGGCCCGCAGAGTTTCAACAAACAGACAGTGGCAAGTTCGAGCAAGCTG 1158  
 QY 341 GlnAlaArgLeuAspAspLeuTrpGluAspIleThrHisSerLeuHisAspGlnGlyHis 360  
 DB 1159 CAGGCCCGCTGATGATACCTGTGGAAAGACATCATCAGCTTCATGACGAGCGCCAC 1218  
 QY 361 SerHisLeuGlyAspPro 366  
 DB 1219 ACCCATCTGGGGAGACCC 1236  
 RESULT 6  
 AAS22623  
 ID AAS22623 standard; cDNA; 1877 BP.  
 AC AAS22623;  
 XX AAS22623;  
 DT 24-OCT-2001 (first entry)  
 XX  
 DE Human cDNA encoding a novel human protein #189.  
 XX  
 KM Human: novel protein; ss; Antinaemic; osteopathic; anti-inflammatory;  
 KM immunomodulatory; cytoskeletal; neuroprotective; vulnerability; nociceptive;  
 KM anticonvulsant; antiallergic; cerebroprotective; antitumoral; antiviral;  
 KM antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;  
 KM thrombolytic; immunogen; antibody; gene therapy; neurological disorder;  
 KM Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;  
 KM tissue regeneration; immune disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200155437-A2.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PE 25-JAN-2001; 2001WO-US02623.  
 XX  
 PR 25-JAN-2000; 2000US-0491404.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Dirmanac RT;  
 XX WPI: 2001-451939/48.  
 DR P-PSDB; AAU14318.  
 XX  
 PT Isolated polypeptides useful for treating anti-inflammatory diseases,  
 PT nervous system disorders, and for regenerating bone and cartilage.  
 XX  
 PS Claim 1; Page 429-431; 894pp; English.  
 CC  
 CC The invention relates to polynucleotides encoding novel human  
 CC proteins or their active domains. The polypeptides, polynucleotides and  
 CC antibodies raised against the polypeptides are used in a method of  
 CC treatment of a mammal and prevention of disorders caused by the aberrant  
 CC protein expression or activity. The polypeptides can be used as  
 CC molecular weight markers, food supplements, and in antibody production.  
 CC The polypeptides are used to identify compounds which bind to the  
 CC polypeptides. Polynucleotides of the invention are used as probes and  
 CC primers, for sequencing, for chromosome or gene mapping, in the  
 CC production of recombinant proteins, and in generating anti-sense DNA or

CC RNA and in gene therapy. Polypeptides of the invention can be used to  
 CC target drugs to a tumour, in assays to determine biological activity, to  
 CC raise antibodies/elicits an immune response, to determine quantitative  
 CC protein levels, as tissue markers, and to isolate receptors or ligands.  
 CC Polypeptides of the invention may also be useful in treating platelet  
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,  
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting  
 CC the proliferation, differentiation and survival of stem cells, as a  
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,  
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral  
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or  
 CC fungal infection or from autoimmunity, cancer, allergy, asthma,  
 CC graft-versus-host disease, eczema, haemophilia, thrombosis,  
 CC anti-inflammatory diseases, nervous system disorders, and infection.  
 CC The present sequence encodes a protein of the invention.  
 XX  
 SQ Sequence 1877 BP; 409 A; 560 C; 584 G; 324 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 0 Length: 1877  
 Score: 366.00 Matches: 366  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 22 Gaps: 0  
 US-09-842-364-3 (1-366) x AAS22623 (1-1877)  
 QY 1 MetAlaSerMetAlaAlaValLeuThrTrpAlaLeuAlaLeuSerAlaPheSerAla 20  
 DB 139 ATGGCAAGCATGGCTGGCGTCTCCTGCTGCTCTTCTTCAAGCTTTTCGGCC 198  
 QY 21 ThrGlnAlaArgLysGlyPheTrpAspTyrPheSerGlnThrSerGlyAspLysGlyArg 40  
 DB 199 ACCCAACACCGAAGAAAGCTTCTCGACATCTTCACCCAGACACCGCGCAAAAGCCAGC 258  
 QY 41 ValGlnGlnIleHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 60  
 DB 259 GTGGAGGAGATCCATTCAGCAGCAAGATGCTGCGAGCCCGGACCTTCMAAGACAGCTT 318  
 QY 61 GlnGlnAspLeuAsnAsnMetAsnLysPheLeuGlnLysLeuAspProLeuSerGlySer 80  
 DB 319 GAGCAAGACCTCAACATATGACAAAGTCTTGAGAAAGCTGAGGCTCTGATGGAGGC 378  
 QY 81 GlnAlaProArgLeuProGlnAspProValGlyMetArgArgGlnGlnGlnGlnGlnGln 100  
 DB 379 GAGCTCTCTCGCGCTCCACAGACCCGCTGCGGACGCGGCGGCGGCGGCGGCGGCGG 438  
 QY 101 GlnGlnValLysAlaArgLeuGlnProTyrMetAlaGlnAlaHisGlnLeuValGlyTyr 120  
 DB 439 GAGAGGCTCAACCTCCCTTCACCTCCACCTTCACAGCCGCGGCGGCGGCGGCGGCTGG 498  
 QY 121 AsnLeuGlnGlyLeuArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 140  
 DB 499 AATTTGAGGAGCTTGGCGGAGCAAGTGAAGCCCTACACGATGATGATGAGGAGAGCTG 558  
 QY 141 AlaLeuArgValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 160  
 DB 559 GCGCTGCGCTCCAGGAGCTGCGAGACAGTGGCTGGCGGGAAGACACCAAGCCG 618  
 QY 161 GlnLeuLeuGlnGlyValLysAlaPheGlnAlaTrpAlaLeuLeuGlnGlnGlnGlnGlnGln 180  
 DB 619 CAGTTGCTGGGGGCGGCGGAGAGCTTGGCTTGGTGGAGAGCTGAGAGCCGCGCTG 678  
 QY 181 ValHisHisThrGlyArgPheLysGlnLeuPheHisProTyrAlaGlnSerLeuValSer 200  
 DB 679 GTGCACACACCGGCGGCTTCAAGAGCTTTCACCCATACCCGAGAGCTGGTGAGC 738  
 QY 201 GlyIleGlyArgHisValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 220  
 DB 739 GGCATCGGGGCGGCGGAGCTGCGAGAGCTGCGAGCTGCTGGCGGAGCCGCGGCGGAGC 798  
 QY 221 ProAlaArgLeuSerArgCysValGlnValLeuSerArgLysLeuThrLeuLysAlaLys 240



```

Db      799  CCCGGGCGCTAGTCGCTGCTGCAGGTCGTCTCTCCCGGAGCTCAGCTCAAGGCCAAG 858
QY      241  AAlaLeuHISAlaArg11eGInGlnAsnLeuAspGlnLeuArg1uGluLeuSerArgAla 260
Db      859  GCCCTGCAGCAGCAGCATCCAGCAAGAACTGAGCAGCAAGCTCGGGAGAGCTCAGCAGAGCC 918
QY      251  PheAlaGlyThrGlyThrGluGluGluValAlaGlyProAspProGlnMetLeuSerGluGlu 280
Db      919  TTTTGACGACCTGGAGCTGACAGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 978
QY      281  ValArg1nArgLeuGlnAlaLapheArg1nAspThrTyLeuGln11eAlaAlaPheThr 300
Db      979  GTGGCCGACGAGCTCAGGCTTCCGCCGAGACCTACCTCAGATACCTGCTTCACT 1038
QY      301  ArgAla11eAspGlnGluThrGluGluValAlaGlnGlnGlnAlaProProProGly 320
Db      1039  CGCGCATCGACGAGAGAGCTGAGAGGTCAGCAGCAGCAGCTGCGCCACCTCCAGCAGCC 1098
QY      321  HISSerAlaPheAlaProGluPheGlnGlnThrAspSerGlyLysValLeuSerLysLeu 340
Db      1099  CACACTGCTCTGCGCCAGAGTTTCAACAACAGCAGAGTGGCAAGCTCTGAGCAAGCTTC 1158
QY      341  GlnAlaArgLeuAspAspLeuTrpGluAsp11eThrHisSerLeuHisAspGlnGlyHis 360
Db      1159  CAGGCCCGCTGTGATGACCTGTGGAGAGCATCATCAGCTTCATCATCAGAGGCCAC 1218
QY      361  SerHisLeuGlyAspPro 366
Db      1219  ACCCATCTGGGGGAGCC 1236

RESULT 7
AAF30036
ID  AAF30036 standard; DNA; 1879 bp.
AC
XX
AC  AAF30036;
XX
DT  23-APR-2001 (first entry)
XX
DE  Human apolipoprotein A-IV-related protein (AA4RP) cDNA.
XX
KW  Apolipoprotein A-IV-related protein; AA4RP; human;
    biallelic marker; lipid metabolism; liver related disorder;
    diagnosis; gene therapy; obesity; diabetes; coronary heart disease;
    ss.
XX
OS  Homo sapiens.
XX
FH  Key
    Location/Qualifiers
    1..20
    FT  5'UTR
    FT  21..1121
    CDS
    FT  /*tag= a
    FT  21..1121
    FT  /*tag= b
    FT  3'UTR
    FT  1122..1879
    FT  /*tag= c
    FT  replace(1153,C,T)
    FT  /*tag= d
    FT  /*note= "biallelic marker 17-42-250, polymorphic
    base C or T"
XX
XX  WO200100803-A2.
XX
XX  04-JAN-2001.
XX
XX  21-JUN-2000; 2000MO-IB01011.
XX
XX  25-JUN-1999; 99US-0141032.
XX  20-DEC-1999; 99MO-IB02058.
XX  21-DEC-1999; 99US-0469099.
XX
XX  (GEST ) GENSET.
XX
PI  Yen F, Denison B, Bour B, Bihain B, Bougueleret L, Duclet A;

```

```

PI  Dumas Mline Edwards J;
XX
XX  WPI: 2001-071485/08.
DR  P-PSDB; AAB20103.
XX
PT  Nucleic acids encoding apolipoprotein A-IV-related proteins (AA4RP) and
PT  biallelic markers of AA4RP, useful for diagnosing lipid metabolism
PT  related disorders and/or liver related disorders.
XX
XX  Claim 5; Page 238-239; 260pp; English.
XX
CC  The present sequence is that of cDNA encoding human apolipoprotein
CC  A-IV-related protein (AA4RP, see AAB20103). AA4RP cDNA clone
CC  117-005-2-0-E10-FLC is deposited as ECACC 99061735. AA4RP is
CC  differentially expressed in obese mouse models, indicating a role
CC  in lipid metabolism disorders. It appears to be the human homologue
CC  of rat regeneration associated protein (RAP3), which is believed to
CC  be involved in liver regeneration, and is likely to have a similar
CC  function. It also shows 52% similarity to apolipoprotein A-IV, and
CC  is likely to have a similar function. The invention also provides
CC  the AA4RP gene, and methods for genotyping a nucleic acid
CC  containing 1 or more of the biallelic markers. Also provided is a
CC  method for detecting a statistical correlation between a biallelic
CC  marker allele and a phenotype. Diagnostic methods are provided for
CC  determining whether an individual is at risk of developing a lipid
CC  metabolism related disorder and/or a liver related disorder, such
CC  as obesity, diabetes and coronary heart disease, and whether the
CC  individual suffers from such a disorder as a result of a
CC  polymorphism in the AA4RP gene. AA4RP DNA constructs can be used
CC  to direct temporal and spatial AA4RP gene expression in recombinant
CC  cell hosts and in transgenic animals. They can also be used for
CC  recombinant production of AA4RP polypeptides, and as probes and
CC  primers.
XX
SQ  Sequence 1879 BP; 395 A; 540 C; 590 G; 353 T; 1 other:

Alignment Scores:
Pred. No.: 0
Score: 366.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 22 Gaps: 0

US-09-842-364-3 (1-366) x AAF30036 (1-1879)
QY  1 MetAlaSerMetAlaValLeuThrTPAlaLeuAlaLeuSerAlaPheSerAla 20
Db  21 ATTCGAACCATGGCTGCCCTGCCCTGACCTGCTCTTTCACCGCTTTTCGCC 80
QY  21 ThrGlnAlaArgLysGlyPheTrpAspTyrPheSerGlnThrSerGlyAspLysGlyArg 40
Db  81 ACCCAGGACGAGAAAGGCTTCTGTGGACTACTTACGACAGCAGCGGGGACAAAGCCAGG 140
QY  41 ValGluGln11eHisGlnGlnLysMetAlaArgGluProAlaThrLeuLysAspSerLeu 60
Db  141 GTGCAGCAGATCATCAGAGAGAGATGCTCGGACCCCGCCGCTGAAAGCACACCTTT 200
QY  61 GluGlnAspLeuAsnAsnMetLysLysPheLeuGluLysLeuArgProLeuSerGlySer 80
Db  201 GAGCAAGACCTCAACAATATGAAAGATTCTCTGAAAAGCTCAGGCTTGAGTGGAGCC 260
QY  81 GluAlaProArgLeuProGlnAspProValAlaGlyMetArgArgGlnLeuGlnGluLeu 100
Db  261 GAGGCTCTCGGCTCCACAGAGACCGGATGCGGAGCTGAGAGGAGGAGTTG 320
QY  101 GluGluValLysAlaArgLeuGlnProTyrMetAlaGlnAlaHisGluLeuValGlyTyr 120
Db  321 GAGGAGGTGAAGGCTGCTCCAGCCCTACATGCGAGAGCCACAGCTGCTGGGCTGG 380
QY  121 AsnLeuGluGlyLeuArgGlnGlnLeuLysProTyrThrMetAspLeuMetGluGlnVal 140

```

```

DB 381 AATTGGAGGCTTGGGAGCACTAAGCCCTACAGATGGATCTGATGGACAGAGTG 440
QY 141 AAlaLeuArgValGlnGluLeuGlnGluGlnLeuArgValValGlyGlyAspThrLysAla 160
DB 441 GCCCTGGCGCTGAGAGAGTGCAGAGACAGTTCGGCTGGGGGGAAGACCAAGGCC 500
QY 161 GlnLeuLeuGlyGlyValAspGluAlaTrpAlaLeuLeuGlnGlyLeuGlnSerArgVal 180
DB 501 CACTTCTGGGGGGCGCGACAGCCCTTGGCTTCTCAGAGCACTGCAGAGCCGCGTG 560
QY 181 ValHisHisThrGlyArgPheLysGlnLeuPheHisProTrpAlaGluSerLeuValSer 200
DB 561 GTGCACCAACACCGCCCTTCAAGAGCTCTTCCACCATACGCGAGAGCTGTGAGC 620
QY 201 GlyIleGlyArgHisValGlnGluLeuHisArgSerValAlaProHisAlaProAlaSer 220
DB 621 GGCATCGGGGCGACGTCGAGAGAGCTGCACCGAGTGTGGCTCCGACGCGCCGACG 680
QY 221 ProAlaArgLeuSerArgCysValGlnValLeuSerArgLysLeuThrLeuLysAlaLys 240
DB 681 CCGCGCGCTCAGTCCCTGCTCCAGTCTCTCCCGAAGCTCAGCTCAAGCCCAAG 740
QY 241 AlaLeuHisAlaArgIleGlnGlnAsnLeuAspGlnLeuArgGlnGluLeuSerArgAla 260
DB 741 GCCCTGCAGCGACGATCCAGCAAGAACCTGACAGCTGGCGCAAGGCTCAGAGAGCC 800
QY 261 PheAlaGlyThrGlyThrGlyGluGlyValGlyProAspProGlnMetLeuSerGlyLys 280
DB 801 TTTCAGGAGCTGAGTGTAGGAAGGGGCGCGCCGAGCCCGCATGCTCTCCGAGAG 860
QY 281 ValArgGlnArgLeuGlnAlaPheArgGlnAspThrLysGlnIleAlaAlaPheThr 300
DB 861 GTCCGCGACGAGCTAGCTGCTTCCGCGAGACACCTTACGATAGCTGCTTCTGACT 920
QY 301 ArgAlaIleAspGlnGluThrGlyGluValGlnGlnGlnLeuAlaProProProGly 320
DB 921 CGGCGCATTCAGCAGAGAGTGCAGAGAGTGCAGAGAGTGGGGCCACCTCCAGCAGGC 980
QY 321 HisSerAlaPheAlaProGluPheGlnGlnThrAspSerGlyLysValLeuSerLysLeu 340
DB 981 CACAGTGCCTTCCGCCAGAGTTTCAACAAACAGACAGTGGCAAGGTTCTGAGCAAGCTG 1040
QY 341 GlnAlaArgLeuAspAspLeuTrpGluAspIleThrHisSerLeuHisAspGlnGlyHis 360
DB 1041 CAGGCCCGCTGATGATGACCTGTGGGAAGACATGATCAGACCTTCTATGACCAAGGCCAC 1100
QY 361 SerHisLeuGlyAspPro 366
DB 1101 AGCCATCTGGGGGAGCCCC 1118

```

RESULT 8  
 AAA87740  
 ID AAA87740 standard; cDNA: 1894 BP.  
 AC AAA87740;  
 XX  
 XX 28-NOV-2000 (first entry)  
 DT  
 XX Human secreted protein encoding cDNA SEQ ID #39.  
 XX  
 XX Human; secreted protein; forensic procedure; gene therapy;  
 KW chromosome mapping; cancer; autoimmune disease; cardiovascular disorder;  
 KW cystic fibrosis; hypothyroidism; immunological disorder; amyloidosis;  
 KW brain disorder; skeletal muscle disorder; eye disorder; obesity;  
 KW mitochondrialcytopathy; diabetes; atherosclerosis; Alzheimer's disease;  
 KW neurodegenerative disorder; graft rejection; dementia; hyperlipidaemia;  
 KW septic shock; impotence; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200037491-A2.  
 XX

```

PD 29-JUN-2000.
XX
XX 20-DEC-1999; 99WO-1802058.
PF
XX 22-DEC-1998; 98US-0113686.
PR
XX 25-JUN-1999; 99US-0141032.
XX
XX (BEST) GENSET.
PA
XX Bougueleret L, Dumas J, Duclert A;
P1
XX
XX WPI: 2000-442637/38.
DR
XX P-PSDB: AAB25778.
PT Polynucleotides and polypeptides encoding proteins with signal
PT peptides, useful in diagnostic, forensic, gene therapy and chromosome
PT mapping procedures -
PS
PS Claim 1; Page 185-187; 306pp; English.
XX
XX This sequence represents human cDNA encoding a secreted protein. The
CC invention relates to sequences AAA87725-A87774 which encode human
CC secreted proteins AAB25763-A25812. The proteins include signal peptides.
CC included in the invention are a host cell containing one of the cDNA
CC sequences, and a purified antibody capable of binding to one of the
CC secreted proteins. Also contained in the invention are methods for
CC storing the sequence data on a computer system, and a method for
CC identifying features of the cDNA sequences using a computer programme.
CC The cDNAs are useful for expressing secreted proteins or fragments to
CC obtain antibodies capable of specifically binding to the secreted
CC proteins. The cDNAs may also be useful in diagnostic, forensic, gene
CC therapy and chromosome mapping procedures and may be used to design
CC expression vectors and secretion vectors. The proteins of the invention
CC may be used to treat diseases including cancer, autoimmune diseases,
CC cardiovascular disorders, cystic fibrosis, hypothyroidism, immunological
CC disorders, amyloidosis, brain disorders, skeletal muscle disorders, eye
CC disorders, obesity, mitochondrialcytopathies, diabetes, atherosclerosis,
CC neurodegenerative disorders, graft rejection, Alzheimer's disease,
CC dementia, hyperlipidaemia, septic shock and impotence.
XX
XX Sequence 1894 BP; 411 A; 541 C; 588 G; 353 T; 1 other;
SQ

```

Alignment Scores:

Pred. No.:	0	Length:	1894
Score:	366.00	Matches:	366
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	21	Gaps:	0

US-09-842-364-3 (1-366) x AAA87740 (1-1894)

```

QY 1 MetaLaserMetalAlaValLeuThrTrpAlaLeuAlaLeuSerAlaPheSerAla 20
DB 21 ATGCAAGCATGGCTGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 80
QY 21 ThrGlnAlaArgLysGlyPheTrpAspPyrPheSerGlnThrSerGlyAspLysGlyArg 40
DB 81 ACCCAGCAGCAAGAAAGCTTCTGGAGTACTTACGACAGACAGCCGAGCAAAAGCAGC 140
QY 41 ValGlnGlnIleHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 60
DB 141 GTGAGCAGATCCCTTCGCGAAGATGCTGCCAGGCCGCGAGCCGAAAGACAGCCTT 200
QY 61 GlnGlnAspLeuAsnAsnMetAsnLysPheLeuGlnLysLeuArgProLeuSerGlySer 80
DB 201 GAGCAAGACCTCAACATATGACAACTTCTGGAAGAAGCTGAGGCTTGTGATGGAGC 260
QY 81 GluAlaProArgLeuProGlnAspProValGlyMetArgArgGlnGlnGlnGlnGlnGln 100
DB 261 GAGGCTCTCGGCTCCACAGAGACCCGCTGGCATCGCGGCGAGCTCAGAGAGATTG 320
QY 101 GlnGluValLysAlaArgLeuGlnProTyrMetAlaGluAlaHisGlnLeuValGlyTyr 120

```

```

Db 321 GAGGAGTGAAGCTCGCTCCAGCCCTACATGGACAGGCGCAGCTGGTGGCTGG 380
QY 121 AsnLeuGluGlyLeuArgGlnGlnLeuLysProTyrThrMetAspLeuMetGlnGlnVal 140
Db 381 AATTGGAGGGCTTGGGAGCAGCAAGTGAAGCCCTACACAGATGATGTGATGGAGAGGTG 440
QY 141 AlaLeuArgValGlnGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 160
Db 441 GCGCTCCGCTGACAGAGCTGACAGAGCACTTCCGCTGGGGAAGCACCAGCCGCC 500
QY 161 GlnLeuLeuGlyGlyValAspGlnAlaTrrPalaLeuLeuGlnGlnGlnGlnGlnGlnGln 180
Db 501 CAGTTCCTGGGGGCGCTGGACAGAGCTTGGGCTTGTCTGAGGAGATGCGAGACCCGCTG 560
QY 181 ValHisHisThrGlyArgPheLysGlyLeuPheHisProTyrAlaCysLeuValSer 200
Db 561 GTGCACCAACACGGCGCTTCAGAAAGAGCTTCCACCCATACCCGAGAGCTGGTAGC 620
QY 201 GlyTLeGlyArgHisValGlnGlnLeuLeuHisArgSerValAlaProHisAlaProAlaSer 220
Db 621 GCGATCGGGCGCCACCTGACAGAGCACTGACAGCTGCTGCTCCGACGCGCCCGCAGC 680
QY 221 ProAlaArgLeuSerArgGlyValGlnValLeuSerArgLysLeuThrLeuLysAlaLys 240
Db 681 CCCGCGCGCTCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 740
QY 241 AlaLeuHisAlaArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260
Db 741 GCGCTGACAGCAGCAGTCCAGCAGCAAGCTGAGCAGCTGCGCAGCAAGCTGAGCAGCAG 800
QY 261 PheAlaGlyThrGlyThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 280
Db 801 TTTTGCAGGAGCTGGAGCTGACAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 860
QY 281 ValArgGlnArgLeuGlnAlaPheArgGlnAspThrTyrLeuGlnGlnGlnGlnGlnGln 300
Db 861 GTTGGCGCAGCAGCTTCCAGCTTCCGCGCAGCAGCAGCTTCCAGCTTCCAGCTTCCAGCT 920
QY 301 ArgAlaIleAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 320
Db 921 CCGCGCATGACAGCAGAGCTGAGGAGGTCCAGCAGCAGCTGCGCGCGCGCGCGCGCGCG 980
QY 321 HisSerAlaPheAlaProGlnPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 340
Db 981 CACAGTCCCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1040
QY 341 GlnAlaArgLeuAspLeuTrrPalaLeuLysProTyrThrMetAspLeuMetGlnGlnVal 360
Db 1041 CAGGCGCGCTGATGACCTTGTGGAGAGCATCTACAGAGCTTGTATGACAGGCGCGCAG 1100
QY 361 SerHisLeuGlyAspPro 366
Db 1101 AGCCATCTGTGGGGAGCCCC 1118

```

RESULT 9  
AAF64022  
ID AAF64022 standard; cDNA; 1894 bp.

AAAF64022:  
05-APR-2001 (first entry)

cdna encoding human secreted protein #23.

Secreted protein; prevention; treatment; diagnosis; disease;

infection; ds.

Homo sapiens.

MO200100806-A2.  
04-JAN-2001.

```

XX 21-JUN-2000; 2000MO-IB00951.
PF 25-JUN-1999; 99US-0141032.
PR 21-DEC-1999; 99US-0469099.
XX (GEST) GENSET.
PA Dumas Milne Edwards J, Houguetier L, Jobert S;
PI MPI; 2001-071487/08.
DR 49 Secreted proteins and the nucleic acids encoding them, useful in
XX gene therapy and for detecting similar sequences in samples.
PT Claim 1; Page 234-235; 307pp; English.
PS The present invention relates to 49 secreted proteins and the cDNAs
XX encoding them. The protein and nucleic acids may be used in the
CC prevention, treatment and diagnosis of diseases associated with
XX inappropriate protein expression.
SQ Sequence 1894 bp; 411 A; 541 C; 588 G; 353 T; 1 other:
Alignment Scores:
Pred. No.: 0 Length: 1894
Score: 366.00 Matches: 366
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0
US-09-842-364-3 (1-366) x AAF64022 (1-1894)
QY 1 MetAlaSerMetAlaIaValIleuThrTrpPalaLeuAlaLeuLeuSerAlaPheSerAla 20
Db 21 ATGCAAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 80
QY 21 ThrGlnAlaArgLysGlyPheThrPaspTyrPheSerGlnThrSerGlyAspLysGlyArg 40
Db 81 ACCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 140
QY 41 ValGlnGlnIleHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 60
Db 141 GTGACACACATTCATCAGCAGCAAGATGCTGCGCAGCCCGCGCAGCCCTGAAGACACAGCTT 200
QY 61 GlnGlnAspLeuAsnAsnMetAsnLysPheLeuGlnLysLeuArgProLeuSerGlySer 80
Db 201 GAGCAAGACCTTCAACATATGAAAGTTCTGGAAGAGCTGAGGCTTGAAGTGGAGAGC 260
QY 81 GluAlaProArgLeuProGlnAspProValGlyMetArgGlnGlnGlnGlnGlnGlnGlnGln 100
Db 261 GAGCCTCTCTCCGCTCCACAGCAGCAGCCGCTGCGCAGCTGCGCAGCAGCAGCAGCTT 320
QY 101 GlnGlnValLysAlaArgLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 120
Db 321 CAGCAGCTCAAGCTTCCCTTCCAGCTTCAATGCGCAGAGCGCAGCAGCTGCTGCGCTGG 380
QY 121 AsnLeuGluGlyLeuArgGlnGlnLeuLysProTyrThrMetAspLeuMetGlnGlnVal 140
Db 381 AATTGGAGGGCTTGGGAGCAGCAAGTGAAGCCCTACAGATGATGTGATGGAGAGGTG 440
QY 141 AlaLeuArgValGlnGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 160
Db 441 GCGCTCCGCTGACAGAGCTGACAGAGCACTTCCGCTGGGGAAGCACCAGCCGCC 500
QY 161 GlnLeuLeuGlyGlyValAspGlnAlaTrrPalaLeuLeuGlnGlnGlnGlnGlnGlnGln 180
Db 501 CAGTTCCTGGGGGCGCTGGACAGAGCTTGGGCTTGTCTGAGGAGATGCGAGACCCGCTG 560
QY 181 ValHisHisThrGlyArgPheLysGlyLeuPheHisProTyrAlaCysLeuValSer 200
Db 561 GTGCACCAACACGGCGCTTCAGAAAGAGCTTCCACCCATACCCGAGAGCTGGTAGC 620

```

QY 201 GYILEGLYARHISYALINGIULLEUHSARSERVALAProHISAlaProHISAlaSer 220  
DB 621 GGCATGGGGCCACCGACGAGCTGCACCGAGTGTGCTCCGACGCGCCCGCCAGC 680  
QY 221 ProHISAlaSerARGCYValGInValLeuSerARGLYSLeuThrLeuLYSAlaLYS 240  
DB 681 CCGGCGGCGCTGATGGCTGCTGCGAGTGTCTCTCCGGAAAGTCAACGCTCAAGGCCAAG 740  
QY 241 AlAlaHISAlaARGLIeGInGInAsnLeuASpGInLeuARGLIuLcUSerARGLa 260  
DB 741 GCCCTCAGCCAGCAGCATCCAGCAACCTGGACACCTGCGCAAGAGCTCAGCAGACC 800  
QY 261 PheAlaGLYThrGLYThrGLYugLyaGLYProAspProGInMeLeuSerGLYGLu 280  
DB 801 TTTCAGGCGCTGCGAGCTGCGAAGGCGCGCCGACCCGACATGCTCTCCGACAGC 860  
QY 281 ValARGLINARGLeugInAlaPheARGLINAsPThrTYrLeuGInLIleAlaAlaPheThr 300  
DB 861 GTGCGCAGGAGCTTACAGGCTTCCGCGACAGACCTACCTGCAGATAGCTGCTTACT 920  
QY 301 ArgAlaIleAspGInGInuThrGInuGInuValGInGInLeuAlaProProProGly 320  
DB 921 CCGGCGATGACGACGAGAGCTGAGAGGTCCACAGCAGCTGCGCACCTCCACAGCAGC 980  
QY 321 HISSerAlaPheAlaProGInuPheGInGInThrAspSerGlyLYSValLeuSerLYSLeu 340  
DB 981 CACAGTGCCTTCCGCCACAGTTTCAACAACAGACAGCTGCGCAAGGTTCTGACCAACCTG 1040  
QY 341 GInAlaARGLeUAsPAsPLeuTrpGInAsPLeuThrHisSerLeuHisAspGInGlyHis 360  
DB 1041 CAGGCCCGTGTGATGACCTGTGGGAAGACATCATCACAGCTTATGACCAAGGCGCAC 1100  
QY 361 SerHisLeuGlyAsPPro 366  
DB 1101 AGCCATCTGGGGGACCCC 1118  
RESULT 10  
AAK89395  
ID AAK89395 standard: DNA; 2249 bp.  
XX  
AC AAK89395;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human digestive system antigen genomic sequence SEQ ID NO: 2971.  
XX  
KM Human; digestive system antigen; gene therapy; cancer; appendicitis;  
KM ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
KM digestive system disorder; Meckel's diverticulum; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200155314-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01324.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214866.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218280.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239335.  
PR 13-OCT-2000; 2000US-0239337.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.

PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249219.  
 PR 17-NOV-2000; 2000US-0249220.  
 PR 17-NOV-2000; 2000US-0249221.  
 PR 17-NOV-2000; 2000US-0249222.  
 PR 17-NOV-2000; 2000US-0249223.  
 PR 17-NOV-2000; 2000US-0249224.  
 PR 17-NOV-2000; 2000US-0249225.  
 PR 17-NOV-2000; 2000US-0249226.  
 PR 17-NOV-2000; 2000US-0249227.  
 PR 17-NOV-2000; 2000US-0249228.  
 PR 17-NOV-2000; 2000US-0249229.  
 PR 17-NOV-2000; 2000US-0249230.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 06-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 XX WPI: 2001-502630/55.  
 DR  
 XX  
 XX Polynucleotides encoding digestive system antigens, useful for  
 PT diagnosing, treating, preventing and/or prognosing disorders of the  
 XX digestive system, particularly cancer and cancer metastases -  
 PS  
 XX Disclosure; SEQ ID NO 2971: 986bp; English.  
 XX  
 XX The present invention provides the protein and coding sequences of a  
 CC number of human digestive system antigens. These can be used in the  
 CC diagnosis, treatment and prevention of digestive system disorders,  
 CC including cancer, Meckel's diverticulum, bacterial or parasitic  
 CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
 CC ulcerative colitis. The present sequence is a genomic DNA fragment  
 CC encoding a digestive system antigen of the invention.  
 XX  
 XX Sequence 2249 BP; 427 A; 671 C; 717 G; 434 T; 0 other;  
 Alignment Scores:

Pred. No.:	3,02e-294	Length:	2249
Score:	312.00	Matches:	312
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	85.25%	Indels:	0
DB:	22	Gaps:	0
US-09-842-364-3 (1-366) x AAK89395 (1-2249)			
QY 55	ThrLeuLysAspSerLeuGluGlnAspLeuAsnMetAsnLysPheLeuGluLysIleu	74	
DB 857	ACCTGAAAGACAGCCTTGAGCAGACCTCAACAAATATGACAGATCTCGGAAAAAGCTG	916	
QY 75	ArgProLeuSerGlySerGluAlaProArgLeuProGlnAspProValGlyMetArgArg	94	
DB 917	AGCCCTCTGACTGCGAGCAGCAGGCTCCCGCTCCACAGACACCCTGGGACATCGCGCGG	976	
QY 95	GlnLeuGlnGluGluLeuGluGluValLysAlaArgLeuGluProTyrMetAlaGluAla	114	
DB 977	CAGCTCAGAGAGAGATTGGAGAGGTGAAGGCTCGCTCCAGCCTTACATGACAGAGCGG	1036	
QY 115	HisGluLeuValGlyTyrPasnLeuGluGlyIleuArgGlnGlnLeuLysProTyrPheMet	134	
DB 1037	CACGAGCTGATGGGCTGGGAATTTGGAGGGCTTGGGACAGCACTGAAGCCCTACAGATG	1096	
QY 135	AspLeuMetGluGlnValAlaLeuArgValGlnGluLeuGlnGluGlnLeuArgValVal	154	
DB 1097	GATCTCATGAGAGAGCTGCGCTCCCGCTCCAGCAGCTCAGAGACAGATTGCGCGCTGTG	1156	
QY 155	GlyGluAspThrLysAlaGlnLeuLeuGlyValAspGluAlaTrpAlaLeuGln	174	
DB 1157	GGGGAAGACACCAAGGCCCATGTTGCTGGGGGGGTGGAGAGGCTTGGCTTGGCTGCGAG	1216	
QY 175	GlyIleuGlnSerArgValAlaHisIsthrcIlyArgPheLysGluLeuPheHisProTyr	194	
DB 1217	GGACTGCAGAGCGCGCTGTGGACACACCGGCGCTTCAAGAGCTCTCCACCCATAC	1276	
QY 195	AlaGluSerLeuValSerGlyIleGlyAlaGlnIleValGlnGluLeuHisArgSerValAla	214	
DB 1277	GCCGAGAGCTGCTGACGCGCATCCGCGCCACGCTCCAGAGCTCAGCCGCACTGTGCT	1336	
QY 215	ProHisAlaProAlaSerProAlaArgLeuSerArgCysValGlnValLeuSerArgLys	234	
DB 1337	CGCGAGCGCCCGCCGAGCGCCCGCCGCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1396	
QY 235	IleuThrLeuLysAlaLysAlaLeuHisAlaArgIleGlnGlnAsnLeuAspGlnLeuArg	254	
DB 1397	CTCAGGCTCAAGGCGCAAGGCTGACACGACGATCCAGACAGACCTGAGCCAGCTGGCC	1456	
QY 255	GluGluLeuSerArgAlaPheAlaGlyThrGlyThrGluGluValAlaGlyProAspPro	274	
DB 1457	GAAGAGCTCAGCAGAGCTTTCAGGCACTCGGACTGAGAGAGGCGCGCGGAGACCC	1516	
QY 275	GlnMetLeuSerGluGluValArgGlnArgLeuGlnAlaPheArgGlnAspThrTyrLeu	294	
DB 1517	CAGATGCTCTCCGAGAGAGGTGGCCAGCGACCTTCCAGGCTTCCGCGAGACACTACTG	1576	
QY 295	GlnIleAlaIaPheThrArgAlaIleAspGlnGluThrGluGluValGlnGlnLeu	314	
DB 1577	CAGATAGCTGCTTCACTCGCCCATGACACGAGACTGAGAGAGGTCCAGAGCAGCTG	1636	
QY 315	AlaProProProGluHisSerAlaPheAlaProGluPheGlnGlnThrAspSerGly	334	
DB 1637	GGCCACCTCCACAGGCGACAGTCTTCCGCCAGAGTTTCAACAAAGACAGTGGC	1696	
QY 335	LysValLeuSerLysLeuGlnAlaArgLeuAspAspLeuTrpGluAspIleThrHisSer	354	
DB 1697	AAGGTTCGAGCAAGCTGAGGCGGCTGTGATGACCTGTGGGAAGACATCACTACAGC	1756	
QY 355	LeuHisAspGlnGlyHisSerHisLeuGlyAspPro	366	
DB 1757	CTTCATACACAGGCGCAGCAGCATCTGGGGAGACCC	1792	

RESULT 11  
AAK89396  
ID AAK89396 standard; DNA; 2250 BP.  
XX  
AC AAK89396;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human digestive system antigen genomic sequence SEQ ID NO: 2972.  
XX  
KW Human; digestive system antigen; gene therapy; cancer; appendicitis;  
ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
digestive system disorder; Meckel's diverticulum; ds.  
XX  
OS Homo sapiens.  
XX  
PN W0200155314-A2.  
XX  
PD 02-AUG-2001.  
XX  
PE 17-JAN-2001; 2001WO-US01324.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 11-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 14-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.

PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235835.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244611.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.

PR 01-DEC-2000: 2000US-0250391.  
 PR 05-DEC-2000: 2000US-0251030.  
 PR 05-DEC-2000: 2000US-0251988.  
 PR 05-DEC-2000: 2000US-0256719.  
 PR 06-DEC-2000: 2000US-0251479.  
 PR 08-DEC-2000: 2000US-0251856.  
 PR 08-DEC-2000: 2000US-0251868.  
 PR 08-DEC-2000: 2000US-0251869.  
 PR 08-DEC-2000: 2000US-0251989.  
 PR 08-DEC-2000: 2000US-0251990.  
 PR 11-DEC-2000: 2000US-0254097.  
 PR 05-JAN-2001: 2001US-02549678.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Roscon CA, Barash SC, Ruben SM;

XX WPI: 2001-502630/55.

XX Polynucleotides encoding digestive system antigens, useful for  
 PT diagnosing, treating, preventing and/or prognosing disorders of the  
 PT digestive system, particularly cancer and cancer metastases -

XX Disclosure: SEQ ID NO 2972; 986bp; English.

XX The present invention provides the protein and coding sequences of a  
 CC number of human digestive system antigens. These can be used in the  
 CC diagnosis, treatment and prevention of digestive system disorders,  
 CC including cancer, Meckel's diverticulum, bacterial or parasitic  
 CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
 CC ulcerative colitis. The present sequence is a genomic DNA fragment  
 CC encoding a digestive system antigen of the invention.

XX Sequence 2250 BP: 427 A; 674 C; 716 G; 433 T; 0 other:

Alignment Scores:  
 Pred. No.: 3,02e-294 Length: 2250  
 Score: 312.00 Matches: 312  
 Percent Similarity: 100.00% Conservative: 0  
 Best local Similarity: 100.00% Mismatches: 0  
 Query Match: 85.25% Indels: 0  
 DB: 22 Gaps: 0

US-09-842-364-3 (1-366) x AAK89396 (1-2250)

QY 55 ThrLeuLysASPserLeuGlnGlnAspLeuAsnMetAsnLysPheLeuGlnLysLeu 74  
 DB 858 ACCCTGAAGAGCAGCTTGAGCAGACCTCAACATATGACAAAGTCTCTGAAAAGCTG 917  
 QY 75 ArgProLeuSerGlySerGlnLysProArgLeuProGlnAspProValGlyMetLysArg 94  
 DB 918 AGGCTCTGAGTGGCAGCAGCGCTCTCGCCCTCCACAGACCGCGTGGCGCTGCGCG 977  
 QY 95 GlnLeuGlnGlnGlnLeuGlnGlnValLysAlaArgLeuGlnProTyrMetAlaLysAla 114  
 DB 978 CAGCTGAGGAGAGAGTGGAGAGGTGAAGGCTCGCTCCAGCCCTACATGACAGAGCG 1037  
 QY 115 HisGlnLeuValGlyTyrAsnLeuGlnGlnLysLeuArgGlnGlnLeuLysProTyrThrMet 134  
 DB 1038 CACGAGCTGGTGGCTGGAATTTGGAGGGCTTGGCGAGCAACTAAGCCCTACACGATG 1097  
 QY 135 AspleuMetGlnGlnValAlaLeuArgValGlnGlnLeuGlnGlnLysArgValAla 154  
 DB 1098 GATCGATGAGAGCAGCTGGCGCTGCGCGTGCAGAGCTGCAGACACTTGGCGCTGCG 1157  
 QY 155 GlyGlnAspTyrLysAlaGlnLeuLeuGlnGlnValLysAlaArgLysAlaLysAla 174  
 DB 1158 GGGGAAGACACCAAGCCCACTTGTGGGGGGCGGAGGAGGCTTGGCTTGTGCGAG 1217  
 QY 175 GlyLeuGlnSerArgValAlaHisHisThrGlyArgPheLysGlnLeuPheHisProTyr 194  
 DB 1218 GGACTGCGAGAGCCGCGTGTGTGCACACACCGCGCTTCANAGAGCTCTTCCACCCATAC 1277

QY 195 AlaGlnSerLeuValSerGlyIleGlyArgHisValGlnGlnLeuHisArgSerValAla 214  
 DB 1278 GCCGAGAGCCTGCTGAGCGGCATCGGGCCCGACGCTGCAGAGCTGCACCGCTGCGCT 1337  
 QY 215 ProHisAlaProHisArgProAlaArgLysSerArgCysValGlnValLeuSerArgLys 234  
 DB 1338 CCGCAGCGCCCGCCAGCGCCCGCGCGCTAGTGGCTGAGGAGGCTCTCCCGAGAG 1397  
 QY 235 LeuThrLeuLysAlaLysAlaLeuHisAlaArgIleGlnGlnAsnLeuAspGlnLeuArg 254  
 DB 1398 CTCAGCCTCAAGGCCAAGGCCCTCGACCGACCGATCGACAGAGAAGCTGCACCTGCC 1457  
 QY 255 GlnGlnLeuSerArgAlaPheAlaGlyThrGlyThrGlnGlnGlnValAlaGlyProAspPro 274  
 DB 1458 GAAGACCTCAGCAGAGCTTTGCGAGGCACTGAGGAGCTGAGAAAGCGCCCGCGACCC 1517  
 QY 275 GlnMetLeuSerGlnGlnValArgGlnArgLysGlnAlaPheArgGlnAspTyrTyrLeu 294  
 DB 1518 CAGATCTCTCCAGAGAGGCGCGCCAGCGACTTCAGGCTTCCGCCAGAGCACTACCTG 1577  
 QY 295 GlnIleAlaAlaPheThrArgAlaIleAspGlnGlnThrGlnGlnGlnValGlnGlnLeu 314  
 DB 1578 CACATAGCTGCTGCTTCACCTCGCGGCATCGACACGACACTGAGGAGCTCCAGCAGCACTG 1637  
 QY 315 AlaProProProProGlnLysSerAlaPheAlaProGlnLysGlnLysThrAspSerGly 334  
 DB 1638 GCCCCACCTCCACACAGCCACGACACTGCTTCCGCCAGACTTTCACAAACAGCAGCTGCG 1697  
 QY 335 LysValLeuSerLysLeuGlnAlaArgLeuAspAspLeuTyrGlnLysLysThrHisSer 354  
 DB 1698 AAGGTTCTAGCAGAGCTGAGAGGCCCTGTGATGACCTGTGGAGAGACATCACTACAGC 1757  
 QY 355 LeuHisAspGlnGlnLysSerHisLeuGlnLysAspPro 366  
 DB 1758 CTTCATGACACAGGCGCCACAGCCTATCTGGGAGACCC 1793

RESULT 12  
 ID AAK89397 standard; DNA; 2250 BP.  
 XX AAK89397;  
 AC  
 XX  
 DT 05-NOV-2001 (first entry)  
 XX  
 DE Human digestive system antigen genomic sequence SEQ ID NO: 2973.  
 XX  
 KW Human; digestive system antigen; gene therapy; cancer; appendicitis;  
 KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
 KW digestive system disorder; Meckel's diverticulum; ds.  
 OS Homo sapiens.  
 PN W0200155314-A2.  
 PD  
 XX  
 XX 02-AUG-2001.  
 PF 17-JAN-2001; 2001WO-0501324.  
 XX  
 PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0188874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 14-AUG-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226868.  
 PR 23-AUG-2000; 2000US-0227182.  
 PR 30-AUG-2000; 2000US-0227009.  
 PR 01-SEP-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 05-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 12-SEP-2000; 2000US-0232081.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 21-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 25-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 26-SEP-2000; 2000US-0234998.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246538.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250191.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 06-DEC-2000; 2000US-0256719.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251858.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 11-JAN-2001; 2000US-0254097.  
 PR 05-JAN-2001; 2000US-0259678.

(HUMA-) HUMAN GENOME SCI INC.  
 Rosen CA, Harash SC, Ruben SM;  
 WPI: 2001-502630/55.

PI Polynucleotides encoding digestive system antigens, useful for  
 PT diagnosing, treating, preventing and/or prognostizing disorders of the  
 PT digestive system, particularly cancer and cancer metastases -  
 PT  
 XX  
 PS Disclosure: SEQ ID NO 2973; 986pp; English.

CC The present invention provides the protein and coding sequences of a  
 CC number of human digestive system antigens. These can be used in the  
 CC diagnosing, treatment and prevention of digestive system disorders,  
 CC including cancer, Meckel's diverticulum, bacterial or parasitic  
 CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
 CC ulcerative colitis. The present sequence is a genomic DNA fragment  
 CC encoding a digestive system antigen of the invention.

XX  
 SQ Sequence 2250 BP; 426 A; 673 C; 718 G; 433 T; 0 other;

Alignment Scores: 3.02e-294 Length: 2250  
 Pred. No.:





FT	/note="17-40..rp"
FT	/tag=
FT	2964..13984
FT	/note="17-41..pu"
FT	complement (34332..14454)
FT	/tag=
FT	u
FT	/note="17-41..rp"
FT	300..318
FT	/tag=
FT	v
FT	/note="17-42-319..mis"
FT	320..338 (42572..42591)
FT	/tag=
FT	w
FT	/note="17-42-319..mis"
FT	3194..3212
FT	/tag=
FT	z
FT	/note="17-41-250..mis"
FT	complement (3214..3232)
FT	/tag=
FT	y
FT	/note="17-41-250..mis"
FT	307..331
FT	/tag=
FT	z
FT	/note="17-42-319 probe"
FT	3201..3225
FT	/tag=
FT	aa
FT	/note="17-41-250 probe"

PN WO200100803-A2.  
XX  
PD 04-JAN-2001.  
XX  
PF 21-JUN-2000; 2000WO-1B01011.  
XX  
PR 25-JUN-1999; 99US-0141032.  
PR 20-DEC-1999; 99WO-1B02058.  
PR 21-DEC-1999; 99US-0469099.  
XX  
XX (GRST ) GENSET.  
PA  
XX  
PI Yen F, Denison B, Bour B, Bihain B, Bouguelerel L, Duclert A;  
PI Dumas Milne Edwards J;  
XX  
XX WPI; 2001-071485/08.  
XR

CC used to direct temporal and spatial AARF gene expression in  
CC recombinant cell hosts and in transgenic animals. They can also  
CC be used for recombinant production of AARF polypeptides, and as

Oy	355	LeuHISASPGINGLYIHLSerHisLcUGLYAspro	366
Dd	3144	CTTCATGACAGGGCCACAGCACACTCGGGAGACC	3179
RESULT 14			
ID	AAF30035		
XX	AAF30035 standard; DNA: 81001 BP.		
XX	AAF30035;		
DT	23-APR-2001	(first entry)	
XX	Human apolipoprotein A-IV-related protein (AA4RP) gene.		
XX			
KW	Apolipoprotein A-IV-related protein; AA4RP; human;		
KW	diabetic marker; lipid metabolism; liver related disorder;		
KW	obesity; diabetes; coronary heart disease; diagnosis; gene therapy;		
XX	chromosome 11: ds.		
XX			
OS	Homo sapiens.		
XX			
Key	Location/Qualifiers		
Ft	5'UTR	10946..12946	
Ft	CDS	/tag= a	
Ft	exon	13478..15210	
Ft	exon	/tag= b	
Ft	intron	12947..12958	
Ft	intron	/tag= c	
Ft	intron	/number= 1	
Ft	exon	12959..13469	
Ft	exon	/tag= d	
Ft	exon	/number= 1	
Ft	intron	13470..13526	
Ft	intron	/tag= e	
Ft	intron	/number= 2	
Ft	exon	13527..13640	
Ft	exon	/tag= f	
Ft	exon	/number= 2	
Ft	exon	13641..13752	
Ft	intron	/tag= g	
Ft	intron	/number= 3	
Ft	intron	13753..14270	
Ft	intron	/tag= h	
Ft	exon	/number= 3	
Ft	exon	14271..15968	
Ft	exon	/tag= i	
Ft	3'UTR	/number= 4	
Ft	3'UTR	15969..17969	
Ft	allele	/tag= j	
Ft	allele	replace(1239,C,T)	
Ft	allele	/tag= k	
Ft	allele	/note= "biallelic marker 20-828-311, polymorphic base C or T"	
Ft	allele	replace(12347,C,T)	
Ft	allele	/tag= l	
Ft	allele	/note= "biallelic marker 17-42-319, polymorphic base C or T"	
Ft	allele	replace(15241,C,T)	
Ft	allele	/tag= m	
Ft	allele	/note= "biallelic marker 17-41-250, polymorphic base C or T"	
Ft	allele	replace(42218,A,G)	
Ft	allele	/tag= n	
Ft	allele	/note= "biallelic marker 20-841-149, polymorphic base A or G"	
Ft	allele	replace(45442,A,G)	
Ft	allele	/tag= o	
Ft	allele	/note= "biallelic marker 20-842-115, polymorphic base A or G"	
Ft	allele	replace(77058,C,T)	
Ft	allele	/tag= p	
Ft	allele	/note= "biallelic marker 20-853-415, polymorphic base A or G"	

FT	primer_bind	929..949	base C or T"
FT		/*tag= q	
FT		/note= "primer 20-828.pu"	
FT	primer_bind	complement (1357..1377)	
FT		/*tag= r	
FT		/note= "primer 20-828.rp"	
FT	primer_bind	12059..12050	
FT		/*tag= s	
FT		/note= "primer 17-42.pu"	
FT	primer_bind	complement (12581..12603)	
FT		/*tag= t	
FT	primer_bind	/note= "primer 17-42.rp"	
FT		14992..15012	
FT		/*tag= u	
FT		/note= "primer 17-41.pu"	
FT	primer_bind	complement (15460..15482)	
FT		/*tag= v	
FT		/note= "primer 17-41.rp"	
FT	primer_bind	42070..42090	
FT		/*tag= w	
FT		/note= "primer 20-841.pu"	
FT	primer_bind	complement (42572..42591)	
FT		/*tag= x	
FT		/note= "primer 20-841.rp"	
FT	primer_bind	45328..45347	
FT		/*tag= y	
FT		/note= "primer 20-842.pu"	
FT	primer_bind	complement (45863..45883)	
FT		/*tag= z	
FT		/note= "primer 20-842.rp"	
FT	primer_bind	76644..76664	
FT		/*tag= aa	
FT		/note= "primer 20-853.pu"	
FT	primer_bind	complement (77166..77185)	
FT		/*tag= ab	
FT		/note= "primer 20-842.rp"	
FT	primer_bind	1220..1238	
FT		/*tag= ac	
FT		/note= "primer 20-828-311.mis"	
FT	primer_bind	complement (1240..1258)	
FT		/*tag= ad	
FT		/note= "primer 20-828-311.mis"	
FT	primer_bind	12328..12346	
FT		/*tag= ae	
FT		/note= "primer 17-42-319.mis"	
FT	primer_bind	complement (12348..12366)	
FT		/*tag= af	
FT		/note= "primer 17-42-319.mis"	
FT	primer_bind	15222..15240	
FT		/*tag= ag	
FT		/note= "primer 17-41-250.mis"	
FT	primer_bind	complement (15242..15260)	
FT		/*tag= ah	
FT		/note= "primer 17-41-250.mis"	
FT	primer_bind	42199..42237	
FT		/*tag= ai	
FT		/note= "primer 20-841-149.mis"	
FT	primer_bind	complement (42219..42237)	
FT		/*tag= aj	
FT		/note= "primer 20-841-149.mis"	
FT	primer_bind	45423..45441	
FT		/*tag= ak	
FT		/note= "primer 20-842-115.mis"	
FT	primer_bind	complement (45443..45461)	
FT		/*tag= al	
FT		/note= "primer 20-842-115.mis"	
FT	primer_bind	77039..77057	
FT		/*tag= am	
FT		/note= "primer 20-853-415.mis"	
FT	primer_bind	complement (77059..77077)	
FT		/*tag= an	
FT		/note= "primer 20-853-415.mis"	





Fri Jan 24 10:09:13 2003

us-09-842-364-3401ig.rng

Page 21

Search completed: January 23, 2003, 18:48:57  
Job time : 289 secs

---

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: January 23, 2003, 17:47:27 (Search time 54 Seconds  
(Without alignments)  
2078.587 Million cell updates/sec)

Title: US-09-842-364-3  
Sequence: 1 MASMAVLTWALALSAFSA.....LWEDITHSLDQSHLGD 366

Scoring table:

OLIGO  
Xgapop 60.0, Ygapext 60.0  
Ygapop 60.0, Ygapext 60.0  
Ygapop 6.0, Ygapext 7.0  
Delop 6.0, Delext 7.0

Searched: 441362 seqs, 153338381 residues

Word size: 6

Total number of hits satisfying chosen parameters: 4253

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame\_plus.p2n.model -DEV=xlh  
-O=/cgn2.1/USPTO.spool/US09842364/runat\_16012003\_164829\_10671/app\_query.fasta.1.519  
-DB=Issued\_Patents\_NA -OPMT=fastlap -SUFFIX=olig.rn1 -MINMATCH=0.1 -LOOPEXT=0  
-LOOPEXT=0 -UNITs=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human4.cdi  
-LIST=45 -DOCALLIN=200 -THR\_SCORE=quality -THR\_MIN=6 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09842364 @cgn2.1 1.25 @runat\_16012003\_164829\_10671 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NCG\_SCORES=0 -WAIT -IONCLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=60 -XGAPEXT=60 -FEAPOP=6 -FEAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database:

Issued\_Patents\_NA.\*  
1: /cgn2.6/prodata/2/ina/5A.COMB.seq.\*  
2: /cgn2.6/prodata/2/ina/5A.COMB.seq.\*  
3: /cgn2.6/prodata/2/ina/6A.COMB.seq.\*  
4: /cgn2.6/prodata/2/ina/6B.COMB.seq.\*  
5: /cgn2.6/prodata/2/ina/6C.COMB.seq.\*  
6: /cgn2.6/prodata/2/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	366	100.0	1879	4	US-09-750-580-2
2	312	85.2	5381	4	US-09-750-580-4
3	312	85.2	81001	4	US-09-750-580-1
4	176	48.1	592	4	US-09-276-531-115
5	48	2.2	48	4	US-08-487-811A-23
6	8	2.2	48	4	US-09-060-694-23
7	8	2.2	105	2	US-08-623-471-6
8	8	2.2	201	2	US-08-623-471-7
9	8	2.2	345	2	US-08-623-471-8
10	8	2.2	397	1	US-08-592-126-106
11	8	2.2	409	1	US-07-626-618A-8
12	8	2.2	409	1	US-08-333-977-8

13	8	2.2	563	1	US-07-928-611-8	Sequence 8, Appl1
14	8	2.2	563	2	US-08-487-811A-8	Sequence 8, Appl1
15	8	2.2	563	4	US-09-060-694-8	Sequence 8, Appl1
16	8	2.2	563	5	US-09-378-074-8	Sequence 8, Appl1
17	8	2.2	659	2	PCT-US93-07370-8	Sequence 10, Appl1
18	8	2.2	659	2	US-07-928-611-10	Sequence 10, Appl1
19	8	2.2	659	2	US-08-487-811A-10	Sequence 10, Appl1
20	8	2.2	659	4	US-09-060-694-10	Sequence 10, Appl1
21	8	2.2	659	4	US-09-378-074-10	Sequence 10, Appl1
22	8	2.2	659	5	PCT-US93-07370-10	Sequence 10, Appl1
23	8	2.2	687	4	US-09-040-984-38	Sequence 38, Appl1
24	8	2.2	687	4	US-09-123-912-38	Sequence 38, Appl1
25	8	2.2	687	4	US-09-643-597-38	Sequence 38, Appl1
26	8	2.2	745	1	US-08-036-5558-163	Sequence 163, App
27	8	2.2	745	1	US-08-469-569-163	Sequence 163, App
28	8	2.2	745	1	US-08-249-322A-163	Sequence 163, App
29	8	2.2	745	1	US-08-469-526A-163	Sequence 163, App
30	8	2.2	745	2	US-08-734-591A-163	Sequence 163, App
31	8	2.2	745	2	US-08-469-660-163	Sequence 163, App
32	8	2.2	745	3	US-08-341-018-51	Sequence 51, Appl
33	8	2.2	745	3	US-08-470-335-163	Sequence 163, App
34	8	2.2	745	4	US-08-735-021-163	Sequence 163, App
35	8	2.2	745	4	US-08-734-664A-163	Sequence 163, App
36	8	2.2	745	4	US-08-470-339-163	Sequence 163, App
37	8	2.2	745	4	US-08-467-602-163	Sequence 163, App
38	8	2.2	745	5	PCT-US94-05083C-159	Sequence 159, App
39	8	2.2	745	5	PCT-US95-0646A-163	Sequence 163, App
40	8	2.2	763	4	US-09-615-192A-117	Sequence 117, App
41	8	2.2	777	2	US-08-975-316-37	Sequence 37, Appl
42	8	2.2	803	1	US-09-615-192A-37	Sequence 37, Appl
43	8	2.2	803	1	US-07-928-611-12	Sequence 12, Appl
44	8	2.2	803	2	US-08-487-811A-12	Sequence 12, Appl
45	8	2.2	803	4	US-09-060-694-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1  
US-09-750-580-2  
Sequence 2, Application US/09750580  
Patent No. 6453280  
GENERAL INFORMATION:  
APPLICANT: Yen, Frances  
APPLICANT: Denison, Blake  
APPLICANT: Bour, Barbara  
APPLICANT: Bihain, Bernard  
APPLICANT: Dumas Mline Edwards, Jean-Baptiste  
APPLICANT: Duclert, Aymeric  
APPLICANT: Bouqueleret, Lydie  
APPLICANT: Ebbels-Reed, Dana  
APPLICANT: Saitor-Cid, Luisa  
TITLE OR INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING NEOPLASTIC CELL GROW  
FILE REFERENCE: 89.052.CIP  
CURRENT APPLICATION NUMBER: US/09750580  
CURRENT FILING DATE: 2000-12-28  
PRIOR APPLICATION NUMBER: US 09/599,362  
PRIOR FILING DATE: 2000-06-21  
PRIOR APPLICATION NUMBER: PCT/IB00/0101  
PRIOR FILING DATE: 2000-06-21  
PRIOR APPLICATION NUMBER: PCT/IB99/02058  
PRIOR FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: US 49/469/099  
PRIOR FILING DATE: 1999-12-21  
PRIOR APPLICATION NUMBER: US 60/113,686  
PRIOR FILING DATE: 1998-12-22  
PRIOR APPLICATION NUMBER: US 60/141,032  
PRIOR FILING DATE: 1999-06-25  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: Patent.pm  
SEQ ID NO 2  
LENGTH: 1879  
TYPE: DNA  
ORGANISM: Homo sapiens

```

FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..29
NAME/KEY: CDS
LOCATION: 30..1121
NAME/KEY: 3'UTR
LOCATION: 1122..1879
NAME/KEY: allele
LOCATION: 1153
OTHER INFORMATION: 17-41-250
US-09-750-580-2 : polymorphic base C or T

Alignment Scores:
Pred. No.: 0 Length: 1879
Score: 366.00 Matches: 366
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

US-09-842-364-3 (1-366) x US-09-750-580-2 (1-1879)

QY 1 MetAlaSerMetAlaAlaValIleuThrTrpAlaLeuAlaLeuSerAlaPheSerAla
21 ATGGCAAGCATGGCTGCGCTGCTCAGCTGGCTGCTCTTTCAGCGTTTCGGCC 80
QY 21 ThrGlnAlaArgLysGlyPheThrPaspTyrPheSerGlnThrSerGlyAspLysGlyArg 40
81 ACCCAAGGACGGAAGGCTTCTGGGACTTCTTCCACCCAGCCAGCGGCGACAAAGCGACG 140
QY 41 ValGlnGlnIleHisGlnGlnIleMetAlaArgLysProAlaThrLeuLysAspSerLeu 60
141 GTGGAGGAGATCCATCCACAGAGATGGCTCGCGAGCCCGCCAGCCCTTGAAGACAGCTT 200
QY 61 GluGlnAspLeuAsnAsnMetAsnLysPheLeuGlnLysLeuArgProLeuSerGlySer 80
201 GAGCAAGACCTCAACATATGAACTTCTGGAAGAGCTGAGGAGGCTCTGAGTGGAGC 260
QY 81 GluAlaProArgLeuProGlnAspProValGlyMetArgArgGlnLeuGlnGlnLeu 100
261 GAGGCTCTCGGCTGCCACAGAGAGCCGGTGGGCAATGGCGGCGGACCTCCAGAGAGCTTG 320
QY 101 GluGlnValLysAlaArgLysGlnProTyrMetAlaGlnAlaHisGlnLeuValGlyTyr 120
321 GAGAGGTGAAGGCTCGGCTCCAGCCCTACATGGCAGAGCGGACAGACTGGTGGCTCG 380
QY 121 AsnLeuGlnGlyLeuArgGlnGlnLeuLysProTyrThrMetAspLeuMetGlnVal 140
381 AATTGGAGGGCTTGGCGGACACTGAAGCCCTACAGATGATCTGATGGAGCAGGTG 440
QY 141 AlaLeuArgValGlnGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 160
441 GCGCTCGCGCGCGGCGGAGAGCTTCCAGCCCTGCGGCGGCGGAGAGAGAGAGAGAGAG 500
QY 161 GlnLeuLeuGlnGlyValAspGlnAlaTrpAlaLeuLeuGlnGlnGlnGlnGlnGln 180
501 CAGTTGCTGGGGGCGGTGGAGCAGAGCTTGGGCTTGGTGGAGGAGCTGAGAGAGAGAG 560
QY 181 ValHisHisThrGlnArgPhePheLysGlnLeuPheHisProTyrAlaGlnSerLeuValSer 200
561 GTGACACACAGCGGCGGCTTAAAGAGCTTCCACCCCTAGCGGAGAGAGAGAGAGAGAG 620
QY 201 GlnTyrLeuGlnArgHisValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 220
621 GCGATCGCGGCGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 680
QY 221 ProAlaArgLeuSerArgCysValGlnValLeuSerArgLysLeuThrLeuLysAlaLys 240
681 CCCCGCGCGGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 740
QY 241 AlaLeuHisAlaArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260
741 GCGCTGACGAGCAGATCCAGAGAACCTGGAGCAGCTGCGGAGAGAGCTCAGCAGAGAGCC 800

```

```

QY 261 PheAlaGlyThrGlyThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 280
DB 801 TTTCAGGACACTGGGAGACTGAGGAAGGCGGCGGCGGAGAGAGAGAGAGAGAGAGAG 860
QY 281 ValArgGlnArgLeuGlnAlaPheArgGlnAspThrTyrLeuGlnIleAlaPheThr 300
DB 861 GTGGCGGAGGACTTCAAGCTTCCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 920
QY 301 ArgAlaAlaAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 320
DB 921 CCGGCGATGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 980
QY 321 HisSerAlaPheAlaProGlnPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 340
DB 981 CACAGTGGCTTGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1040
QY 341 GlnAlaArgLeuAspAspLeuProGlnAspPheThrHisSerLeuHisAspGlnGln 360
DB 1041 CAGGCGGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1100
QY 361 SerHisLeuGlyAspPro 366
DB 1101 AGCCATCTGGGGAGAGCC 1118

RESULT 2
US-09-750-580-4
Sequence 4, Application US/09750580
Patent No. 6455280
GENERAL INFORMATION:
APPLICANT: Yen, Frances
APPLICANT: Denison, Blake
APPLICANT: Bour, Barbara
APPLICANT: Bihain, Bernard
APPLICANT: Dumas, Milne Edwards, Jean-Baptiste
APPLICANT: Duchetel, Aymeric
APPLICANT: Bouquelierel, Lydie
APPLICANT: Ebbels-Reed, Dana
APPLICANT: Salter-Cid, Luisa
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING NEOPLASTIC CELL GROWTH
FILE REFERENCE: 89,US2,CIP
CURRENT FILING DATE: US/09/750,580
CURRENT FILING DATE: 2000-12-28
PRIORITY FILING DATE: US 09/599,362
PRIORITY FILING DATE: 2000-06-21
PRIORITY FILING DATE: PCT/IB00/0101
PRIORITY FILING DATE: 2000-06-21
PRIORITY FILING DATE: PCT/IB99/02058
PRIORITY FILING DATE: 1999-12-20
PRIORITY FILING DATE: US 49/469/099
PRIORITY FILING DATE: 1999-12-21
PRIORITY FILING DATE: US 60/113,686
PRIORITY FILING DATE: 1998-12-22
PRIORITY FILING DATE: 1998-12-22
PRIORITY FILING DATE: 1999-06-25
PRIORITY FILING DATE: 1999-06-25
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Patent.pm
SEQ ID NO 4
LENGTH: 5381
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..918
OTHER INFORMATION: 5' regulatory region
NAME/KEY: exon
LOCATION: 919..930
OTHER INFORMATION: exon 1
NAME/KEY: exon
LOCATION: 1442..1498
OTHER INFORMATION: exon 2
NAME/KEY: exon
LOCATION: 1613..1724

```



```

OTHER INFORMATION: exon 3
NAME/KEY: exon
LOCATION: 2243..3940
OTHER INFORMATION: exon 4
NAME/KEY: misc_feature
LOCATION: 3941..5381
OTHER INFORMATION: 3'regulatory region
NAME/KEY: allele
LOCATION: 319
OTHER INFORMATION: 17-42-319 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 3213
OTHER INFORMATION: 17-41-250 : polymorphic base C or T
NAME/KEY: conflict
LOCATION: 1241
OTHER INFORMATION: 17-39-343 : T in ref genbank AC007707
NAME/KEY: conflict
LOCATION: 1447
OTHER INFORMATION: 17-40-202 : G in ref genbank AC007707
NAME/KEY: primer_bind
LOCATION: 1..11022
OTHER INFORMATION: 17-42-.rp complement
NAME/KEY: primer_bind
LOCATION: 553..11575
OTHER INFORMATION: 17-42-.rp complement
NAME/KEY: primer_bind
LOCATION: 899..11920
OTHER INFORMATION: 17-39-.pu
NAME/KEY: primer_bind
LOCATION: 1246..12267
OTHER INFORMATION: 17-40-.pu
NAME/KEY: primer_bind
LOCATION: 1441..12461
OTHER INFORMATION: 17-39-.rp complement
NAME/KEY: primer_bind
LOCATION: 1632..12651
OTHER INFORMATION: 17-40-.rp complement
NAME/KEY: primer_bind
LOCATION: 2964..13984
OTHER INFORMATION: 17-41-.pu
NAME/KEY: primer_bind
LOCATION: 3432..14454
OTHER INFORMATION: 17-41-.rp complement
NAME/KEY: primer_bind
LOCATION: 300..318
OTHER INFORMATION: 17-42-319.mis
NAME/KEY: primer_bind
LOCATION: 320..338
OTHER INFORMATION: 17-42-319.mis complement
NAME/KEY: primer_bind
LOCATION: 3194..3212
OTHER INFORMATION: 17-41-250.mis
NAME/KEY: primer_bind
LOCATION: 3214..3232
OTHER INFORMATION: 17-41-250.mis complement
NAME/KEY: misc_binding
LOCATION: 307..331
OTHER INFORMATION: 17-42-319.probe
NAME/KEY: misc_binding
LOCATION: 3201..3225
OTHER INFORMATION: 17-41-250.probe
US-09-750-580-4

```

## Alignment Scores:

```

Pred. No.: 4.3e-305      Length: 5381
Score: 312.00           Matches: 312
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 85.25%             Indels: 0
DB: 4                     Gaps: 0

```

US-09-842-364-3 (1-366) x US-09-750-580-4 (1-5381)

```

Oy 55 ThrleuLysAspSerLeuGlnAspLeuAsnMeLanLysPheLeuGluLysLeu 74
Db 2244 ACCCTGAAGACAGCCTTGACCAAGACCTCAACATATGACCAAGCTTCTGAAAAGCTG 2303
Oy 75 ArgProLeuSerGlySerGlnAlaProArgLeuProGlnAspProValGlyMetLArg 94
Db 2304 AGGCTCTGAGTGGAGAGGAGGCTCTGCGCCACAGACGCCGCTGGATCGCGCCG 2363
Oy 95 GlnLeuGlnGluGluLeuGlnGluValLysAlaArgLeuGlnProTyrMetAlaGlnAla 114
Db 2364 CAGCTGACAGGAGGATTGGAGAGGTGAAGGCTCCCTCCAGCCCTACATGACAGAGCG 2423
Oy 115 HisGlnLeuValGlyTyrPasnLeuGlnGlyLeuArgGlnGlnLysProTyrThiMet 134
Db 2424 CACGAGCTGGTGGCTGGATTTGAGGGCTTGGGGACGACACTGAAACCTTACACGATG 2483
Oy 135 AspLeuMetGlnGlnValAlaLeuArgValGlnGlnLeuGlnGlnValAla 154
Db 2484 GATCTGATGAGACAGAGTGGCCCTGCGGCTGCAGACGCTGCAGACAGCTTCGCGCTG 2543
Oy 155 GlyLysAspThrLysAlaGlnLeuLeuGlyGlyValAspGlnAlaTrrAlaLeuGln 174
Db 2544 GGGAGAGACACCAAGGCCAGCTTCTGGGGGCGTGGACGAGGCTTGGCTTCTGCGAG 2603
Oy 175 GlyLeuGlnSerArgValAlaHisThrGlyArgPheLysGlyLeuPheHisProTyr 194
Db 2604 GGACTGCAGAGCCCGGTGGTCCACACACCGCGCTTCAAGAGCTTCTCCACCCATAC 2663
Oy 195 AlaGlnSerLeuValSerGlyLeuGlyArgHisValGlnGlnLeuHisArgSerValAla 214
Db 2664 GCCGAGACCCCTGTCAGCGGCGCATGCGGCGCCAGCGACGAGCTGACCCGAGCTGGCT 2723
Oy 215 ProHisAlaProAlaSerProAlaArgLeuSerArgCysValGlnValLeuSerArgLys 234
Db 2724 CCGCAGCCGCCCGCCAGCCCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2783
Oy 235 LeuThrLeuLysAlaLysAlaLeuHisAlaArgGlnGlnGlnAsnLeuAspGlnLeuArg 254
Db 2784 CTCAGGCTCAAGGCCAAGGCCCTGACGACGACGACATCCAGCAAGAACTTGGACAGCTG 2843
Oy 255 GlnGlnLeuSerAlaArgAlaPheAlaGlyThrGlyThrGlnGlnAlaGlyProAspPro 274
Db 2844 GAAGAGCTTCAGCAGAGCTTTCAGCCACTGCGACGCTGCGAAGGCGCCCGCCGACCC 2903
Oy 275 GlnMetLeuSerGlnGlnValArgGlnArgGlnArgGlnAlaPheArgGlnAspThrLys 294
Db 2904 CAGATGCTCTCCGAGGAGGTGCGCAGCGACTTCAGGCTTTCGCGCAGGACACTTACTG 2963
Oy 295 GlnIleAlaIlePheThrArgAlaIleAspGlnGlnArgGlnGlnGlnGlnGlnGln 314
Db 2964 CAAATAGCTGCTTCACTCCGCGCATGACGACGAGACTGAGAGGTCCAGCAGCAGCTG 3023
Oy 315 AlaProProProProGlnHisSerAlaPheAlaProGlnPheGlnGlnIntrAspSerGly 334
Db 3024 CCGCCACCTCCACACGCCCGCAGAGTGCCTTCGCCCAAGATTTCACCAAGACAGACTG 3083
Oy 335 LysValLeuSerLysLeuGlnAlaArgLeuAspAspLeuTrpGlnAspIleThrHisSer 354
Db 3084 AAGGCTTGAGCAAGAGTGGAGGCCGCTGATGATACCTGTGGAGACATCACTACAGC 3143
Oy 355 LeuHisAspGlnGlnHisSerHisLeuGlyAspPro 366
Db 3144 CTCATGACCAAGGCCACAGCCTTGGGGGACCC 3179

```

## RESULT 3

US-09-750-580-1  
Sequence 1, Application US/09750580  
Patent No. 6455280

## GENERAL INFORMATION:

APPLICANT: Yen, Frances  
APPLICANT: Denison, Blake  
APPLICANT: Bour, Barbara  
APPLICANT: Bihain, Bernard

APPLICANT: Dumas Milne Edwards, Jean-Baptiste  
APPLICANT: Duclert, Aymeric  
APPLICANT: Bouquelard, Lydie  
APPLICANT: Eddels-Reed, Dana  
APPLICANT: Saller-Cid, Luisa  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING NEOPLASTIC CELL GROWTH  
FILE REFERENCE: 89.052.CIP  
CURRENT APPLICATION NUMBER: US/09/750,580  
PRIORITY FILING DATE: 2000-12-28  
PRIORITY FILING DATE: 2000-06-21  
PRIORITY FILING DATE: 2000-06-21  
PRIORITY FILING DATE: 2000-06-21  
PRIORITY FILING DATE: 2000-06-21  
PRIORITY FILING DATE: 1999-12-20  
PRIORITY FILING DATE: 1999-12-20  
PRIORITY FILING DATE: 1999-12-21  
PRIORITY FILING DATE: 1998-12-22  
PRIORITY FILING DATE: 1998-12-22  
PRIORITY FILING DATE: 1999-06-25  
SOFTWARE: PatentL.pm  
SEQ ID NO: 1  
LENGTH: 81001  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 10946..12946  
OTHER INFORMATION: 5' regulatory region  
NAME/KEY: exon  
LOCATION: 12947..12958  
OTHER INFORMATION: exon 1  
NAME/KEY: exon  
LOCATION: 13470..13526  
OTHER INFORMATION: exon 2  
NAME/KEY: exon  
LOCATION: 13641..13752  
OTHER INFORMATION: exon 3  
NAME/KEY: exon  
LOCATION: 14271..15968  
OTHER INFORMATION: exon 4  
NAME/KEY: misc.feature  
LOCATION: 15969..17969  
OTHER INFORMATION: 3' regulatory region  
NAME/KEY: allele  
LOCATION: 1239  
OTHER INFORMATION: 20-828-311 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 12347  
OTHER INFORMATION: 17-42-319 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 15241  
OTHER INFORMATION: 17-41-250 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 42218  
OTHER INFORMATION: 20-841-149 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 45442  
OTHER INFORMATION: 20-842-115 : polymorphic base A or C  
NAME/KEY: allele  
LOCATION: 77058  
OTHER INFORMATION: 20-853-415 : polymorphic base C or T  
NAME/KEY: primer\_bind  
LOCATION: 929..949  
OTHER INFORMATION: 20-828-pu  
NAME/KEY: primer\_bind  
LOCATION: 1357..1377  
OTHER INFORMATION: 20-828-rp complement  
NAME/KEY: primer\_bind  
LOCATION: 12029..12050  
OTHER INFORMATION: 17-42-pu

NAME/KEY: primer\_bind  
LOCATION: 12581..12603  
OTHER INFORMATION: 17-42-rp complement  
NAME/KEY: primer\_bind  
LOCATION: 14992..15012  
OTHER INFORMATION: 17-41-pu  
NAME/KEY: primer\_bind  
LOCATION: 15460..15482  
OTHER INFORMATION: 17-41-rp complement  
NAME/KEY: primer\_bind  
LOCATION: 42070..42090  
OTHER INFORMATION: 20-841-pu  
NAME/KEY: primer\_bind  
LOCATION: 42572..42591  
OTHER INFORMATION: 20-841-rp complement  
NAME/KEY: primer\_bind  
LOCATION: 45328..45347  
OTHER INFORMATION: 20-842-pu  
NAME/KEY: primer\_bind  
LOCATION: 45863..45883  
OTHER INFORMATION: 20-842-rp complement  
NAME/KEY: primer\_bind  
LOCATION: 76644..76664  
OTHER INFORMATION: 20-853-pu  
NAME/KEY: primer\_bind  
LOCATION: 77166..77185  
OTHER INFORMATION: 20-853-rp complement  
NAME/KEY: primer\_bind  
LOCATION: 1220..1238  
OTHER INFORMATION: 20-828-311.mis  
NAME/KEY: primer\_bind  
LOCATION: 1240..1258  
OTHER INFORMATION: 20-828-311.mis complement  
NAME/KEY: primer\_bind  
LOCATION: 12328..12346  
OTHER INFORMATION: 17-42-319.mis  
NAME/KEY: primer\_bind  
LOCATION: 12348..12366  
OTHER INFORMATION: 17-42-319.mis complement  
NAME/KEY: primer\_bind  
LOCATION: 15222..15240  
OTHER INFORMATION: 17-41-250.mis  
NAME/KEY: primer\_bind  
LOCATION: 15242..15260  
OTHER INFORMATION: 17-41-250.mis complement  
NAME/KEY: primer\_bind  
LOCATION: 42199..42217  
OTHER INFORMATION: 20-841-149.mis  
NAME/KEY: primer\_bind  
LOCATION: 42219..42237  
OTHER INFORMATION: 20-841-149.mis complement  
NAME/KEY: primer\_bind  
LOCATION: 45423..45441  
OTHER INFORMATION: 20-842-115.mis  
NAME/KEY: primer\_bind  
LOCATION: 45443..45461  
OTHER INFORMATION: 20-842-115.mis complement  
NAME/KEY: primer\_bind  
LOCATION: 77039..77057  
OTHER INFORMATION: 20-853-415.mis  
NAME/KEY: primer\_bind  
LOCATION: 77059..77077  
OTHER INFORMATION: 20-853-415.mis complement  
NAME/KEY: primer\_bind  
LOCATION: 1227..1251  
OTHER INFORMATION: 20-828-311-probe  
NAME/KEY: misc.binding  
LOCATION: 12335..12359  
OTHER INFORMATION: 17-42-319-probe  
NAME/KEY: misc.binding  
LOCATION: 15229..15253  
OTHER INFORMATION: 17-41-250-probe  
NAME/KEY: misc.binding

LOCATION: 42206..42230  
 OTHER INFORMATION: 20-841-149.probe  
 NAME/KEY: misc.binding  
 LOCATION: 45430..45454  
 OTHER INFORMATION: 20-842-115.probe  
 NAME/KEY: misc.binding  
 LOCATION: 77046..77070  
 OTHER INFORMATION: 20-853-415.probe  
 US-09-750-580-1

## Alignment Scores:

Prod. No.:	6.45e-304	Length:	81001
Score:	312.00	Matches:	312
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	85.25%	Indels:	0
DB:	4	Gaps:	0

US-09-842-364-3 (1-366) x US-09-750-580-1 (1-81001)

QY 55 ThrLeuLysAspSerLeuGluGlnAspLeuAsnMetLysPheLeuGluLysLeu 74  
 DB 14272 ACCCTGAAGACACACCTTGAGCAAGACCTCAACAATATGAACAAGTCTTGAAAAAGCTG 14331  
 QY 75 ArgProLeuSerClySerGlyAlaIayArgLeuProGlnAspProValGlyMetArg 94  
 DB 14332 AGGCTCTGAGTGGAGGAGGAGGCTCTCGGCTCCACAGAGCCCGGTGGCATGCGGCG 14391  
 QY 95 GlnLeuGlnGluGluGluGluGluValLysAlaArgLeuGlnProTyrMetAlaGluAla 114  
 DB 14392 CAGCTGAG 14451  
 QY 115 HisGluLeuValGlyTyrPheLeuGluGlyLeuArgGlnGlnLeuLysProTyrThrMet 134  
 DB 14452 CAGAGAGTGGTGGGCTGGAATTTGGAGGCTTGCGGAGCAGACCTGAAGCCCTAACGATG 14511  
 QY 135 AspleuMetClyGlnValAlaLeuArgValGlnGluLeuGlnGlnGlnGlnGlnGlnGln 154  
 DB 14512 GATCTGTGAG 14571  
 QY 155 GlnGluAspThrLysAlaGlnLeuGluGlyValAspGluAlaTrrPheAlaLeuGln 174  
 DB 14572 GGGGAGACACCAAGGCGCGAGTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14631  
 QY 175 GlyLeuGlnSerArgValAlaHisHisThrGlyArgPheLysGluLeuPheHisProTyr 194  
 DB 14632 GGAAGTGGAGAGCGCGCGTGGAGACACCGCGCGCTTCAAGAGCTCTTCCACCCATAC 14691  
 QY 195 AlagluSerLeuValSerGlyIleGlyArgHisValGlnGluLeuHisArgSerValAla 214  
 DB 14692 GCGGAGAGCTGTGAGCGGCGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14751  
 QY 215 ProHisAlaProLysSerProAlaArgLysSerArgCysValGlnValLeuSerArgLys 234  
 DB 14752 CCGCAGCCCGCCCGCAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14811  
 QY 235 LeuThrLeuLysAlaLysAlaLeuHisAlaArgIleGlnGlnAsnLeuAspGlnLeuArg 254  
 DB 14812 CTCAGCGCTCAAGGCGCAAGGCGCTGACGACGATCCAGAGAACTGGAGACACTGCGC 14871  
 QY 255 GlnGluLeuSerArgAlaPheAlaGlyThrGlyThrGlnGluGluGlyAlaGlyProAspPro 274  
 DB 14872 GAAGAGCTGAGCAGAGCTTGTGAGGAGCTGAGAGAGGAGGCGCGCGCGCGCGCGCG 14931  
 QY 275 GlnMetLeuSerGluGluValArgGlnArgLeuGlnAlaPheArgGlnAspThrLysLeu 294  
 DB 14932 CAGATGCTCTCCAGAGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14991  
 QY 295 GlnIleAlaAlaPheThrArgAlaIleAspGlnGluThrGlnGluValGlnGlnGlnLeu 314  
 DB 14992 CAGATGCTCTCTCACTCG 15051  
 QY 315 AlaProProProProGlyHisSerAlaPheAlaProGluPheGlnGlnThrAspSerGly 334

DB 15052 CCGCAGCTCCAGCAGCG 15111  
 QY 335 LysValLeuSerLysLeuGlnAlaArgLeuAspAspLeuTrrPheAlaIleThrHisSer 354  
 DB 15112 AAGCTTGTGACCAAGCTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15171  
 QY 355 LeuHisAspGlnGlnLysHisSerHisLeuGluAspPro 366  
 DB 15172 CTCATGAGCAGGCG 15207

## RESULT 4

US-09-276-531-115  
 Sequence 115, Application US/09276531  
 Patent No. 6183968

## GENERAL INFORMATION:

APPLICANT: Bandman, Olga  
 APPLICANT: Lei, Preethi  
 APPLICANT: Hillman, Jennifer L.  
 APPLICANT: Yuc, Henry

APPLICANT: Guebler, Karl J.  
 APPLICANT: Baughn, Mariah R.

TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING  
 NUMBER OF INVENTIONS: RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
 STREET: 3174 PORTER DRIVE  
 CITY: PALO ALTO  
 STATE: CALIFORNIA  
 COUNTRY: USA  
 ZIP: 94304

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/276,531  
 FILING DATE: Herewith

CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 60/079,677  
 FILING DATE: March 27, 1998

CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Lynn E. Murty, Ph.D.  
 REGISTRATION NUMBER: 42,918  
 REFERENCE/DOCKET NUMBER: PA-0008 US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (650) 855-0555  
 TELEFAX: (650) 845-4166

INFORMATION FOR SEQ ID NO: 115:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 592 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear

IMMEDIATE SOURCE:  
 LIBRARY: L1VHTUT04  
 CLONE: 2512827

US-09-276-531-115

Alignment Scores:  
 Prod. No.: 9.9e-169  
 Score: 176.00  
 Percent Similarity: 100.00%  
 Best Local Similarity: 100.00%  
 Query Match: 48.09%

Prod. No.:	9.9e-169	Length:	592
Score:	176.00	Matches:	176
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	48.09%	Indels:	0
DB:	4	Gaps:	0

US-09-842-364-3 (1-366) x US-09-276-531-115 (1-592)

```

Qy 1 MetAlaSerMetAlaAlaValLeuThrPrAlaLeuAlaLeuSerAlaPheSerAla 20
Db 35 ATGGCAAGCATGGCTGCCCTGCCTCAGCTGGCTTCTTTTACGGCTTTTCGGCC 94
Qy 21 ThrGlnAlaArgLysGlyPheTrpAspTyrPheSerGlnThrSerGlyAspLysGlyArg 40
Db 95 ACCGAGCAGCAGAAAGAGCTTCTGGAGTACTTCCAGCAGACAGCGGGAGCAAGAGCAGG 154
Qy 41 ValGlnGlnIleHISGlnGlnLysMetAlaArgGluProAlaThrLeuLysAspSerLeu 60
Db 155 GTGAGCAGATCCATCGACGAAAGATGGCTCGCAGCCCGCAGCCCTGAAAGACAGCCTT 214
Qy 61 GluGlnAspLeuAsnAsnMetAsnLysPheLeuGlnLysLeuArgProLeuSerGlySer 80
Db 215 GAGCAAGACCTCAACATATGACAACAACTTCTGAAAGAGCTGAGCCCTTCAGTGGAGCC 274
Qy 81 GluAlaProArgLeuProGlnAspProValGlyMetArgArgGlnLeuGlnGluLeu 100
Db 275 GAGGCTCCTCGGCTCCACAGACCCCGTGCGCATGCGCGGAGCTGCAGAGAGAGAGTTG 334
Qy 101 GluGlnValLysAlaArgLeuGlnProTyrMetAlaGlnAlaHisGlnLeuValGlyTyr 120
Db 335 GAGGAGGTGAAGGCTCGCTCCAGCCCTACATGGCAGAGCGCAGAGCTGGTGGCTGG 394
Qy 121 AsnLeuGlnGlyLeuArgGlnGlnLeuLysProTyrThrMetAspLeuMetGlnGlnVal 140
Db 395 AATTTGAGCGGCTTCCGCGCAGCACTGAACCCCTACACAGATGATCTGATGAGACAGCTG 454
Qy 141 AlaLeuArgValGlnGlnLeuGlnGlnGlnLeuAlaArgValAlaGlyLysPheLysAla 160
Db 455 GCCCTCGGCTGAGGAGGCTGAGGAGCAGTCCGCTGTGGGGAGACACCAAGCC 514
Qy 161 GlnLeuLeuGlnGlyValAlaAspGlnAlaTrpAlaLeuLeuGlnGlnGln 176
Db 515 CAGTCTCGGGGGGCGGTGGAGCAGGCTTGGGCTTGTGTCAGAGGAGCTG 562

RESULT 5
US-08-487-811A-23
; Sequence 23, Application US/08487811A
; Patent No. 5883226
; GENERAL INFORMATION:
; APPLICANT: Civeilli, Olivier
; TITLE OF INVENTION: A No. 5883226el Human Dopamine Receptor and Uses
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boenhen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,811A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5883226nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 90,1092-L
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; TELEX:
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; Alignment Scores:
; Pred. No.: 3.43
; Score: 8.00
; Percent Similarity: 100.00%
; Best Local Similarity: 100.00%
; Query Match: 2.19%
; DB: 2
; Gaps: 0
US-09-842-364-3 (1-366) x US-08-487-811A-23 (1-48)
Qy 82 AlaProArgLeuProGlnAspPro 89
Db 5 GCCGCCGCGCTCCCGCAGAGCCCGC 28

RESULT 6
US-09-060-694-23
; Sequence 23, Application US/09060694
; Patent No. 6203998
; GENERAL INFORMATION:
; APPLICANT: Civeilli, Olivier
; TITLE OF INVENTION: A No. 6203998el Human Dopamine Receptor and Uses
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boenhen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/060,694
; FILING DATE: 15-APR-1998
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6203998nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 90,1092-MM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; TELEX:
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; Alignment Scores:
; Pred. No.: 3.43
; Score: 8.00
; Percent Similarity: 100.00%
; Best Local Similarity: 100.00%
; Query Match: 2.19%
; DB: 4
; Gaps: 0
US-09-842-364-3 (1-366) x US-09-060-694-23 (1-48)
Qy 82 AlaProArgLeuProGlnAspPro 89

```

```
|||||
DB      5 GCGCCCCGCTCCCGACGACCC 28

RESULT 7
US-08-623-471-6
; Sequence 6, Application US/08623471
; Patent No. 5846823
; GENERAL INFORMATION:
; APPLICANT: Allelix Biopharmaceuticals Inc
; APPLICANT: Owolabi, Joshua
; APPLICANT: Rampersad, Vikarna
; APPLICANT: Kamboj, Rajender
; TITLE OF INVENTION: STABLE D4 CELL LINES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allelix Biopharmaceuticals Inc
; STREET: 6850 Goreway Drive
; CITY: Mississauga
; COUNTRY: Canada
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/623,471
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA94/00538
; FILING DATE: 27-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: RIDOUT & MAYBEE, Attn. Robert G. Hiron
; REFERENCE/DOCKET NUMBER: ALLEL/51B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416)-868-1482
; TELEFAX: (416)-362-0823
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-08-623-471-6

Alignment Scores:
Pred. No.: 7.49      Length: 105
Score: 8.00         Matches: 8
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 2.19%      Indels: 0
DB: 2               Gaps: 0

US-09-842-364-3 (1-366) x US-08-623-471-6 (1-105)

OY      82 AlaprogLeuprogInaspro 89
DB      10 GCGCCCCGCTCCCGACGACCC 33

RESULT 8
US-08-623-471-7
; Sequence 7, Application US/08623471
; Patent No. 5846823
; GENERAL INFORMATION:
; APPLICANT: Allelix Biopharmaceuticals Inc
; APPLICANT: Owolabi, Joshua
; APPLICANT: Rampersad, Vikarna
; APPLICANT: Kamboj, Rajender
; TITLE OF INVENTION: STABLE D4 CELL LINES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allelix Biopharmaceuticals Inc
; STREET: 6850 Goreway Drive
; CITY: Mississauga
; COUNTRY: Canada
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/623,471
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA94/00538
; FILING DATE: 27-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: RIDOUT & MAYBEE, Attn. Robert G. Hiron
; REFERENCE/DOCKET NUMBER: ALLEL/51B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416)-868-1482
; TELEFAX: (416)-362-0823
; TELEX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-08-623-471-7

Alignment Scores:
Pred. No.: 14.3     Length: 201
Score: 8.00         Matches: 8
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 2.19%      Indels: 0
DB: 2               Gaps: 0

US-09-842-364-3 (1-366) x US-08-623-471-7 (1-201)

OY      82 AlaprogLeuprogInaspro 89
DB      10 GCGCCCCGCTCCCGACGACCC 33

RESULT 9
US-08-623-471-8
; Sequence 8, Application US/08623471
; Patent No. 5846823
; GENERAL INFORMATION:
; APPLICANT: Allelix Biopharmaceuticals Inc
; APPLICANT: Owolabi, Joshua
; APPLICANT: Rampersad, Vikarna
; APPLICANT: Kamboj, Rajender
; TITLE OF INVENTION: STABLE D4 CELL LINES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allelix Biopharmaceuticals Inc
; STREET: 6850 Goreway Drive
; CITY: Mississauga
; COUNTRY: Canada
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/623,471
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA94/00538
; FILING DATE: 27-SEP-1994
; ATTORNEY/AGENT INFORMATION:
```

NAME: RIDOUT & MAYBEE, Attn. Robert C. Hiron  
REFERENCE/DOCKET NUMBER: ALLFL/518  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416)-868-1482  
TELEFAX: (416)-362-0823  
TRIEX:  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 345 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-623-471-8

Alignment Scores:  
Pred. No.: 24.6 Length: 345  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.19% Indels: 0  
DB: Gaps: 0

US-09-842-364-3 (1-366) x US-08-623-471-8 (1-345)

QY 82 AlaProArgLeuProGlnAspPro 89  
Db 10 GCGCCCGCCTCCCGCAGAGACCC 33

RESULT 10  
US-08-592-126-106/c  
Sequence 106, Application US/08592126  
Patent No. 5821091  
GENERAL INFORMATION:  
APPLICANT: Gregory Dolganov  
TITLE OF INVENTION: Transcripts Encoding Immunomodulatory  
TITLE OF INVENTION: polypeptides  
NUMBER OF SEQUENCES: 151  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dellinger & Associates  
STREET: 350 Cambridge Avenue, Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/592,126  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Sholtz, Charles K.  
REGISTRATION NUMBER: 38,615  
REFERENCE/DOCKET NUMBER: 4600-0111  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 106:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 397 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: G127.seq  
US-08-592-126-106

Alignment Scores:  
Pred. No.: 28.2 Length: 397  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.19% Indels: 0  
DB: Gaps: 0

US-09-842-364-3 (1-366) x US-08-592-126-106 (1-397)

QY 217 AlaProAlaSerProAlaArgLeu 224  
Db 140 GCTCCAGCCAGCCGCGCAGACTT 117

## RESULT 11

US-07-626-618A 8

Sequence 8, Application US/07626618A  
Patent No. 5422265

GENERAL INFORMATION:

APPLICANT: Van Tol, Hubert H.M.

APPLICANT: Cleveli, Olivier

TITLE OF INVENTION: A No. 5422265el Human Dopamine Receptor and Uses

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: Allegretti & Witcoff, Ltd.

STREET: 10 South Wacker Drive, Suite 3000

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/626,618A

FILING DATE: 7 DEC 1990

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: No. 5422265nan, Kevin E

REGISTRATION NUMBER: 35,303

REFERENCE/DOCKET NUMBER: 90,1092

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-715-1000

TELEFAX: 312-715-1234

TELEX: 810-221-8317

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 409 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: exon

LOCATION: 1..409

FEATURE:

NAME/KEY: CDS

LOCATION: 2..409

US-07-626-618A-8

## Alignment Scores:

Pred. No.: 29.1

Score: 8.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 2.19%

DB: 1

Length: 409  
Matches: 8  
Conservative: 0  
Mismatches: 0  
Indels: 0  
Gaps: 0

US-09-842-364-3 (1-366) x US-07-626-618A-8 (1-409)

QY 82 AlaProArgLeuProGlnAspPro 89

```

|||||
DB      350 GCGCCCGCTCCCGCAGAGACCC 373

RESULT 12
US-08-333-977-8
; Sequence 8, Application US/08333977
; Patent No. 5594108
; GENERAL INFORMATION:
;   APPLICANT: Van Tol, Hubert H.M.
;   TITLE OF INVENTION: A No. 5594108el Human Dopamine Receptor and Uses
;   NUMBER OF SEQUENCES: 22
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Allegretti & Wilcoff, Ltd.
;   STREET: 10 South Wacker Drive, Suite 3000
;   CITY: Chicago
;   STATE: Illinois
;   COUNTRY: USA
;   ZIP: 60606
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patent Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/333,977
;   FILING DATE: 03-NOV-1994
;   CLASSIFICATION: 530
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/626,618
;   FILING DATE: 7 DEC 1990
;   ATTORNEY/AGENT INFORMATION:
;   NAME: No. 5594108man, Kevin E
;   REGISTRATION NUMBER: 35,303
;   REFERENCE/DOCKET NUMBER: 90,1092
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 312-715-1000
;   TELEFAX: 312-715-1234
;   TELEX: 810-221-8317
;   INFORMATION FOR SEQ ID NO: 8:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 409 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 1..409
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 2..409
; US-08-333-977-8

Alignment Scores:
Pred. No.:      29.1      Length:      409
Score:          8.00      Matches:      8
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match:    2.19%      Indels:      0
                        Gaps:      0
US-09-842-364-3 (1-366) x US-08-333-977-8 (1-409)

OY      82 AlaProArgLeuProGlnAspPro 89
DB      350 GCGCCCGCTCCCGCAGAGACCC 373

RESULT 13
US-07-928-611-8
; Sequence 8, Application US/07928611
; Patent No. 5569601
; GENERAL INFORMATION:
;   APPLICANT: Van Tol, Hubert H.M.
;   TITLE OF INVENTION: A No. 5569601el Human Dopamine Receptor and Uses
;   NUMBER OF SEQUENCES: 22
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Allegretti & Wilcoff, Ltd.
;   STREET: 10 South Wacker Drive, Suite 3000
;   CITY: Chicago
;   STATE: Illinois
;   COUNTRY: USA
;   ZIP: 60606
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patent Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/928,611
;   FILING DATE: 19920810
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: No. 5569601man, Kevin E
;   REGISTRATION NUMBER: 35,303
;   REFERENCE/DOCKET NUMBER: 90,1092-B
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 312-715-1000
;   TELEFAX: 312-715-1234
;   TELEX: 810-221-8317
;   INFORMATION FOR SEQ ID NO: 8:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 563 base pairs
;   TYPE: NUCLEIC ACID
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 1..563
;   IDENTIFICATION METHOD: experimental
;   OTHER INFORMATION: /evidence= EXPERIMENTAL
;   OTHER INFORMATION: /standard_name= "Alternate Exon 3: D4.2"
;   OTHER INFORMATION: /note= "This sequence represent the sequence of
;   the third exon of allele D4.2 of the human D4
;   dopamine receptor gene"
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 257..262
;   IDENTIFICATION METHOD: experimental
;   OTHER INFORMATION: /function= "Polymorphic PstI site"
;   OTHER INFORMATION: /evidence= EXPERIMENTAL
;   OTHER INFORMATION: /label= PstI
;   OTHER INFORMATION: /note= "this feature is the site of one of the
;   restriction enzymes whereby digestion of genomic
;   DNA produces a RFLP "
;   FEATURE:
;   NAME/KEY: repeat_region
;   LOCATION: 346..442
;   OTHER INFORMATION: /rpt_type= "random"
;   OTHER INFORMATION: /rpt_unit= 348 .. 396
;   OTHER INFORMATION: /note= "This sequence represents one of 7 known
;   alleles of human D4 dopamine receptor gene
;   encoding a 16 amino acid sequence repeated twice"
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 2..563
; US-07-928-611-8

Alignment Scores:
Pred. No.:      40      Length:      563
Score:          8.00      Matches:      8
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match:    2.19%      Indels:      0

```



DB: 1 Gaps: 0

US-09-842-364-3 (1-366) x US-07-928-611-8 (1-563)

QY 82 AlaprogaleuprogInaspro 89  
 |||  
 Db 350 GCGCCCGCTCCCGAGACCC 373

RESULT 14

US-08-487-811A-8  
 ; Sequence 8, Application US/08487811A  
 ; Patent No. 5883226  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Civeilli, Olivier  
 ; APPLICANT: Van Tol, Hubert H.M.  
 ; TITLE OF INVENTION: A No. 5883226e1 Human Dopamine Receptor and Uses  
 ; NUMBER OF SEQUENCES: 24  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff  
 ; STREET: 300 South Wacker Drive  
 ; CITY: Chicago  
 ; STATE: IL  
 ; COUNTRY: USA  
 ; ZIP: 60606  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/487,811A  
 ; FILING DATE: 07-JUN-1995  
 ; CLASSIFICATION: 530  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: No. 5883226nan, Kevin E  
 ; REGISTRATION NUMBER: 35,303  
 ; REFERENCE/DOCKET NUMBER: 90,1092-L  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312-913-0001  
 ; TELEFAX: 312-913-0002  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 8:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 563 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: exon  
 ; LOCATION: 1..563  
 ; IDENTIFICATION METHOD: experimental  
 ; OTHER INFORMATION: /evidence= EXPERIMENTAL  
 ; OTHER INFORMATION: /standard\_name= "Alternative Exon 3: D4.2"  
 ; OTHER INFORMATION: /note= "This sequence represent the sequence of  
 ; the third exon of allele D4.2 of the human D4  
 ; OTHER INFORMATION: dopamine receptor gene"  
 ; FEATURE:  
 ; NAME/KEY: misc.feature  
 ; LOCATION: 257..262  
 ; IDENTIFICATION METHOD: experimental  
 ; OTHER INFORMATION: /function= "Polymorphic PstI site"  
 ; OTHER INFORMATION: /evidence= EXPERIMENTAL  
 ; OTHER INFORMATION: /label= PstI  
 ; OTHER INFORMATION: /note= "This feature is the site of one of the  
 ; restriction enzymes whereby digestion of genomic  
 ; OTHER INFORMATION: DNA produces a RFLP "  
 ; FEATURE:  
 ; NAME/KEY: repeat\_region  
 ; LOCATION: 346..442  
 ; OTHER INFORMATION: /rpt\_type= "tandem"  
 ; OTHER INFORMATION: /rpt\_unit= 348..396  
 ; OTHER INFORMATION: /note= "This sequence represents one of 7 known

OTHER INFORMATION: alleles of human D4 dopamine receptor gene  
 ; FEATURE: encoding a 16 amino acid sequence repeated twice  
 ; NAME/KEY: CDS  
 ; LOCATION: 2..563

US-08-487-811A-8

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
8.00	100.00%	563	8
Best Local Similarity:	100.00%		Conservative: 0
Query Match:	2.19%		Mismatches: 0
DB:	2		Indels: 0
			Gaps: 0

US-09-842-364-3 (1-366) x US-08-487-811A-8 (1-563)

QY 82 AlaprogaleuprogInaspro 89  
 |||  
 Db 350 GCGCCCGCTCCCGAGACCC 373

RESULT 15

US-09-060-694-8  
 ; Sequence 8, Application US/09060694  
 ; Patent No. 6203998  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Civeilli, Olivier  
 ; APPLICANT: Van Tol, Hubert H.M.  
 ; TITLE OF INVENTION: A No. 6203998e1 Human Dopamine Receptor and Uses  
 ; NUMBER OF SEQUENCES: 24  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff  
 ; STREET: 300 South Wacker Drive  
 ; CITY: Chicago  
 ; STATE: IL  
 ; COUNTRY: USA  
 ; ZIP: 60606  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/060,694  
 ; FILING DATE: 15-APR-1998  
 ; CLASSIFICATION: 530  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: No. 6203998nan, Kevin E  
 ; REGISTRATION NUMBER: 35,303  
 ; REFERENCE/DOCKET NUMBER: 90,1092-XM  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312-913-0001  
 ; TELEFAX: 312-913-0002  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 8:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 563 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: exon  
 ; LOCATION: 1..563  
 ; IDENTIFICATION METHOD: experimental  
 ; OTHER INFORMATION: /evidence= EXPERIMENTAL  
 ; OTHER INFORMATION: /standard\_name= "Alternative Exon 3: D4.2"  
 ; OTHER INFORMATION: /note= "This sequence represent the sequence of  
 ; the third exon of allele D4.2 of the human D4  
 ; OTHER INFORMATION: dopamine receptor gene"  
 ; FEATURE:  
 ; NAME/KEY: misc.feature  
 ; LOCATION: 257..262

```

: IDENTIFICATION METHOD: experimental
: OTHER INFORMATION: /function= "Polymorphic PstI site"
: OTHER INFORMATION: /evidence= EXPERIMENTAL
: OTHER INFORMATION: /label= PstI
: OTHER INFORMATION: /note= "This feature is the site of one of the
: OTHER INFORMATION: restriction enzymes whereby digestion of genomic
: OTHER INFORMATION: DNA produces a RFLP "
: FEATURE:
: NAME/KEY: repeat_region
: LOCATION: 346..442
: OTHER INFORMATION: /rpt_type= "tandem"
: OTHER INFORMATION: /rpt_unit= 348..396
: OTHER INFORMATION: /note= "This sequence represents one of 7 known
: OTHER INFORMATION: alleles of human D4 dopamine receptor gene
: OTHER INFORMATION: encoding a 16 amino acid sequence repeated twice
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 2..563
: US-09-060-694-8

```

```

Alignment Scores:
Pred. No.: 40          Length: 563
Score: 8.00          Matches: 8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%    Mismatches: 0
Query Match: 2.19%              Indels: 0
DB: 4                          Gaps: 0

```

US-09-842-364-3 (1-366) x US-09-060-694-8 (1-563)

```

OY  82 AlaProArgLeuProGlnAspPro 89
    |||||||||||||||||||||
Db  350 GCGCCCGCGCCTCCCGCAGAGCCCC 373

```

Search completed: January 23, 2003, 19:24:03  
Job time : 71 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 23, 2003, 17:48:17 : Search time 67 Seconds  
(without alignments)  
2454.193 Million cell updates/sec

Title: US-09-842-364-3

Sequence: 1 NASMAVLTWALTLAFAFA.....LMEDITHSLHDOGHSLGDP 366

Scoring table:

OLIGO  
Xgapop 60.0, Xgapext 60.0  
Ygapop 60.0, Ygapext 60.0  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 396772 seqs, 224632407 residues

Word size: 6

Total number of hits satisfying chosen parameters: 6466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-O=/cgn2\_1/USPTO.SP001/US09842364/runat.16012003.164830.10683/app.query.fasta.1.519  
-DB=Published.Applications\_NA\_QCWT-lastap\_SUFFIX-olig.rnpb\_MINMATCH-0.1  
-LOOPEXT-0 -ONITS-bits -START-1 -END-1 -MATRIX-oligo  
-TRANS-human40.ccl -LIST-45 -DOCALIGN-200 -THR.SCORE-quality -THR.MIN-6  
-ALIGN-15 -MODE-LOCAL -OUTFMT-pto -NORM-ext -HEAPSIZ-500 -MINLEN-0  
-MAXLEN-200000000 -USER-US09842364.@CGN.1.1.33@runat.16012003.164830.10683  
-NCPU-6 -ICPU-3 -NO.XLPXY -NO.MMAP -LARGEQUERY -NEG.SCORES-0 -WAIT -LONGILOC  
-DEV.TIMEOUT-120 -WARN.TIMEOUT-30 -THREDS-1 -XGAPOP-60 -XGAPEXT-60 -Fgapop-6  
-Fgapext-7 -Ygapop-60 -Ygapext-60 -DELop-6 -DELext-7

Database: Published.Applications\_NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCR\_NEM\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEM\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US07\_NEM\_PUB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCR\_NEM\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEM\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEM\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEM\_PUB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEM\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	366	100.0	1858	US-09-835-996A-1
2	366	100.0	1877	US-09-835-996A-10
3	366	100.0	1879	US-09-751-877-2
4	312	85.2	5381	US-09-751-877-4

5	312	85.2	81001	US-09-751-877-1	Sequence 1, Appl 1
6	279	76.2	1393	US-09-800-729-40	Sequence 40, Appl 1
7	213	58.2	1334	US-09-349-015-1	Sequence 1, Appl 1
8	70	19.1	468	US-09-800-729-209	Sequence 209, Appl 1
9	33	9.0	331	US-09-800-729-210	Sequence 210, Appl 1
10	9	2.5	301	US-09-983-965-4674	Sequence 4674, Appl 1
11	9	2.5	397	US-09-983-965-4360	Sequence 4360, Appl 1
12	12	2.5	405	US-09-960-352-10825	Sequence 10825, Appl 1
13	13	2.5	419	US-09-960-352-8211	Sequence 8211, Appl 1
14	14	2.5	832	US-09-938-842-2586	Sequence 2586, Appl 1
15	9	2.5	2850	US-09-815-242-7104	Sequence 7104, Appl 1
16	8	2.2	158	US-09-796-692-3304	Sequence 3304, Appl 1
17	8	2.2	368	US-09-983-965-4711	Sequence 4711, Appl 1
18	8	2.2	426	US-09-983-965-5038	Sequence 5038, Appl 1
19	8	2.2	445	US-09-864-761-2633	Sequence 2633, Appl 1
20	8	2.2	505	US-09-783-590-242	Sequence 242, Appl 1
21	8	2.2	522	US-09-864-761-32402	Sequence 32402, Appl 1
22	8	2.2	582	US-09-864-761-15898	Sequence 15898, Appl 1
23	8	2.2	687	US-09-735-705-38	Sequence 38, Appl 1
24	8	2.2	687	US-09-850-716A-38	Sequence 38, Appl 1
25	8	2.2	687	US-09-897-778-38	Sequence 38, Appl 1
26	5	2.2	755	US-09-809-545A-9	Sequence 9, Appl 1
27	8	2.2	948	US-09-886-055-176	Sequence 176, Appl 1
28	8	2.2	1006	US-09-946-807-104	Sequence 104, Appl 1
29	8	2.2	1006	US-09-795-686-104	Sequence 104, Appl 1
30	8	2.2	1006	US-09-795-686-104	Sequence 104, Appl 1
31	8	2.2	1413	US-09-938-842A-422	Sequence 422, Appl 1
32	8	2.2	1458	US-09-741-669-106	Sequence 106, Appl 1
33	8	2.2	1532	US-09-992-598-150	Sequence 150, Appl 1
34	8	2.2	1532	US-09-989-293A-150	Sequence 150, Appl 1
35	8	2.2	1532	US-09-989-735-150	Sequence 150, Appl 1
36	8	2.2	1532	US-09-990-44A-150	Sequence 150, Appl 1
37	8	2.2	1532	US-09-989-730-150	Sequence 150, Appl 1
38	8	2.2	1532	US-09-990-436-150	Sequence 150, Appl 1
39	8	2.2	1532	US-09-991-181-150	Sequence 150, Appl 1
40	8	2.2	1532	US-09-993-687-150	Sequence 150, Appl 1
41	8	2.2	1532	US-09-989-735-150	Sequence 150, Appl 1
42	8	2.2	1532	US-09-997-653-150	Sequence 150, Appl 1
43	8	2.2	1532	US-10-174-590-121	Sequence 121, Appl 1
44	8	2.2	1532	US-10-176-758-121	Sequence 121, Appl 1
45	8	2.2	1532	US-10-175-737-121	Sequence 121, Appl 1

#### ALIGNMENTS

RESULT 1  
US-09-835-996A-1  
Sequence 1, Application US/09835996A  
Patent No. US20020142953A1  
GENERAL INFORMATION:  
APPLICANT: Baillinger, Dennis  
APPLICANT: Loeb, Debra  
APPLICANT: Montgomery, Julie  
APPLICANT: Tang, Y. Tom  
APPLICANT: Zhou, Ping  
APPLICANT: Goodrich, Ryle  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhao, Qing  
APPLICANT: Wehrman, Tom  
APPLICANT: Drmanac, Radoje  
APPLICANT: Ren, Feiyan  
APPLICANT: Qian, Xiaohong  
APPLICANT: Wang, Dunru  
TITLE OF INVENTION: MATERIALS AND METHODS RELATING TO LIPID METABOLISM  
FILE REFERENCE: 28110/35915A  
CURRENT APPLICATION NUMBER: US/09/835, 996A  
CURRENT FILING DATE: 2001-04-16  
PRIOR APPLICATION NUMBER: US 60/197,137  
PRIOR FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: US 09/714,936  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: US 09/667,298





Db 201 GAGCAAGACCTCAACATATGACAACTGCTGAGAAAGCTGAGGCTCTGATGGGAGC 260  
 QY 81 GUAAlaProArgLeuProGlnAspProValGlyMetArgArgGlnLeuGlnGluLeu 100  
 Db 261 GAGGCTCTCTGCTCCACAGACCGGCTGGGCAWGGCGGCGAGCTGCAGAGAGACTTG 320  
 QY 101 GUGUValAlaAlaArgLeuGlnProTyrMetAlaGlnAlaHisGlnLeuValGlyTyr 120  
 Db 321 GAGGAGGTGAAGGCTGCTCCCTCCAGACCTACATGGCAGAGCGGCGAGCTGGCTGC 380  
 QY 121 AsnLeuGluGlyLeuArgGlnGlnLeuGlySerTyrThrMetAspLeuMetGluGlnVal 140  
 Db 381 AATTGGAGGCTTGGCGGAGCAACTGAACTGACACGATGATGATGATGATGAGAGAGGTG 440  
 QY 141 AlaLeuArgValGlnGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 160  
 Db 441 GCGCTGCGCTGCGAGCGCTGCAGAGAGAGCTGCGCTGCGCGGCGAGAGAGAGAGAG 500  
 QY 161 GlnLeuLeuGlyGlyValAlaAspGlnAlaTyrAlaLeuLeuGlnGlnGlnGlnGln 180  
 Db 501 CAGTTGCTGGGGGGCGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 560  
 QY 181 ValHisHisThrGlyArgGlyLeuGlyLeuGlyLeuGlyLeuGlyLeuGlyLeuGly 200  
 Db 561 GTGACACACACCGCGCGCTTCAAGAGAGCTCTCCACCCATACGCGAGAGAGAGAGAG 620  
 QY 201 GilylGlyArgHisValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 220  
 Db 621 GCGATCGCGCGCGCGAG 680  
 QY 221 ProAlaArgLeuSerArgGlyValGlnValLeuSerArgGlyLeuThrLeuValAla 240  
 Db 681 CCGCGCGCGCTCAAGTCCGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 740  
 QY 241 AlaLeuHisAlaArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260  
 Db 741 GCGCTGACAG 800  
 QY 261 PheAlaGlyThrGlyThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 280  
 Db 801 TTTGCAAGGCACTGGAG 860  
 QY 281 ValArgGlnArgLeuGlnAlaPheArgGlnAspThrTyrLeuGlnGlnGlnGlnGln 300  
 Db 861 GTGCGCGAG 920  
 QY 301 ArgAlaAlaAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 320  
 Db 921 CCGGCGAG 980  
 QY 321 HisSerAlaPheAlaProGlnPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 340  
 Db 981 CACAGTACCTTGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1040  
 QY 341 GlnAlaArgLeuAspAspLeuTyrGlnAspLeuThrHisSerLeuHisAspGlnGln 360  
 Db 1041 CAGCGCGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1100  
 QY 361 SerHisLeuGlyAspPro 366  
 Db 1101 AGCCATCTGGGGAG 1118  
 RESULT 4  
 US-09-751-877-4  
 Sequence 4, Application US/09751877  
 Patent No. US20020142949A1  
 GENERAL INFORMATION:  
 APPLICANT: Ten, Frances  
 APPLICANT: Denison, Blake  
 APPLICANT: Bour, Barbara  
 APPLICANT: Bihain, Bernard  
 APPLICANT: Dumas Milne Edwards, Jean-Baptiste  
 APPLICANT: Duclert, Aymeric

APPLICANT: Bougueleret, Lydie  
 APPLICANT: Ebbels-Reed, Dana  
 APPLICANT: Salter-Cid, Luisa  
 TITLE OR INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING NEOPLASTIC CELL GROWTH  
 FILE REFERENCE: 89.US3.REG  
 CURRENT APPLICATION NUMBER: US/09/751,877  
 NUMBER OF SEQ ID NOS: 6  
 SOFTWARE: Patent.pm  
 SEQ ID NO 4  
 LENGTH: 5381  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: 1..918  
 OTHER INFORMATION: 5'regulatory region  
 NAME/KEY: exon  
 LOCATION: 919..930  
 OTHER INFORMATION: exon 1  
 NAME/KEY: exon  
 LOCATION: 1442..1498  
 OTHER INFORMATION: exon 2  
 NAME/KEY: exon  
 LOCATION: 1613..1724  
 OTHER INFORMATION: exon 3  
 NAME/KEY: exon  
 LOCATION: 2243..3940  
 OTHER INFORMATION: exon 4  
 NAME/KEY: misc\_feature  
 LOCATION: 3941..5381  
 OTHER INFORMATION: 3'regulatory region  
 NAME/KEY: allele  
 LOCATION: 319  
 OTHER INFORMATION: 17-42-319 : polymorphic base C or T  
 NAME/KEY: allele  
 LOCATION: 3213  
 OTHER INFORMATION: 17-41-250 : polymorphic base C or T  
 NAME/KEY: conflict  
 LOCATION: 1241  
 OTHER INFORMATION: 17-39-343 : T in ref genbank AC007707  
 NAME/KEY: conflict  
 LOCATION: 1447  
 OTHER INFORMATION: 17-40-202 : G in ref genbank AC007707  
 NAME/KEY: primer\_bind  
 LOCATION: 1..11022  
 OTHER INFORMATION: 17-42-pu  
 NAME/KEY: primer\_bind  
 LOCATION: 553..11575  
 OTHER INFORMATION: 17-42-rp complement  
 NAME/KEY: primer\_bind  
 LOCATION: 899..11920  
 OTHER INFORMATION: 17-39-pu  
 NAME/KEY: primer\_bind  
 LOCATION: 1246..12267  
 OTHER INFORMATION: 17-40-pu  
 NAME/KEY: primer\_bind  
 LOCATION: 1441..12461  
 OTHER INFORMATION: 17-39-rp complement  
 NAME/KEY: primer\_bind  
 LOCATION: 1632..12651  
 OTHER INFORMATION: 17-40-rp complement  
 NAME/KEY: primer\_bind  
 LOCATION: 2964..13984  
 OTHER INFORMATION: 17-41-pu  
 NAME/KEY: primer\_bind  
 LOCATION: 3432..14454  
 OTHER INFORMATION: 17-41-rp complement  
 NAME/KEY: primer\_bind  
 LOCATION: 300..318  
 OTHER INFORMATION: 17-42-319 mis  
 NAME/KEY: primer\_bind  
 LOCATION: 320..338

OTHER INFORMATION: 17-42-319.mis complement  
 NAME/KEY: primer\_bind  
 LOCATION: 3194..3212  
 OTHER INFORMATION: 17-41-250.mis  
 NAME/KEY: primer\_bind  
 LOCATION: 3214..3232  
 OTHER INFORMATION: 17-41-250.mis complement  
 NAME/KEY: misc\_binding  
 LOCATION: 307..331  
 OTHER INFORMATION: 17-42-319.probe  
 NAME/KEY: misc\_binding  
 LOCATION: 3201..3225  
 OTHER INFORMATION: 17-41-250.probe  
 US-09-751-877-4

Alignment Scores:  
 Pred. No.: 7.62e-297 Length: 5381  
 Score: 312.00 Matches: 312  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 85.25% Indels: 0  
 DB: 10 Caps: 0

US-09-842-364-3 (1-366) x US-09-751-877-4 (1-5381)

QY 55 ThrLeuLysAspSerLeuGluGlnAspLeuAsnMetAsnLysPheLeuGluLysLeu 74  
 DB 2244 ACCCTGAAACACACACCTTGACCAAGACCTACACATATGCAACAGTTCTCGAAGAACCTG 2303  
 QY 75 ArgProLeuSerGlySerGluAlaProArgLeuProGlnAspProValGlyMetLargArg 94  
 DB 2304 AGCCCTCTGAGTGAGGAGGAGGCTCCCTCCAGACAGACCCGGTGGGATCGGCGG 2363  
 QY 95 GlnLeuGlnGluGluGluGluValLysAlaArgLeuGlnProTyrMetAlaGluAla 114  
 DB 2364 CAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2423  
 QY 115 HisGluLeuValGlyTyrPheLeuGluGlyLeuArgGlnGlnLeuLysProTyrThMet 134  
 DB 2424 CACGACCTGCTGGGCTGGGATTTGCAAGCCCTTCGCCGACGACGACGACGACGACGACG 2483  
 QY 135 AspLeuMetGluGlnValAlaLeuArgValGlnGluLeuGlnGlnGlnGlnGlnGlnGln 154  
 DB 2484 GATCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2543  
 QY 155 GlyGluAspThrLysAlaGlnLeuLeuGlyValAlaAspGluAlaTyrPheLeuGln 174  
 DB 2544 GGGGAAAGACACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2603  
 QY 175 GlyLeuGlnSerArgValAlaHisHisThrGlyArgPheLysGluLeuPheHisProTyr 194  
 DB 2604 GCACCTGCAGAGCCCGGCTGGCAGCACCCGCCCTTCGCAAGACGCTTCGCAAGCCGATAC 2663  
 QY 195 AlaGluSerLeuValSerGlyIleGlyArgHisValGlnGluLeuHisAspSerValAla 214  
 DB 2664 GCCGAAGAGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2723  
 QY 215 ProHisAlaProLysSerProAlaArgLeuSerArgCysValGlnValLeuSerArgLys 234  
 DB 2724 CGGACAGGCGCCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2783  
 QY 235 LeuThrLeuLysAlaLysAlaLeuHisAlaArgIleGlnGlnAsnLeuAspGlnLeuArg 254  
 DB 2784 CTCACCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2843  
 QY 255 GluGluLeuSerArgAlaPheAlaGlyThrGlyThrGluGluLysAlaGlyProAspPro 274  
 DB 2844 GAAGAGCTCAGCAGGAGCTTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2903  
 QY 275 GlnMetLeuSerGluGluValAlaArgGlnArgLeuGlnAlaPheArgGlnAspThrTyrLeu 294  
 DB 2904 CAGATGCTCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2963

QY 295 GlnIleAlaAlaPheThrArgAlaIleAspGlnGluThrGluGluValGlnGlnGlnLeu 314  
 DB 2964 CAGATAGCTGCTTACCTACCTCGGCGCCATCGACAGGAGGAGGAGGAGGAGGAGGAGGAGG 3023  
 QY 315 AlaProProProProGlyHisSerAlaPheAlaProGluPheGlnGlnThrAspSerGly 334  
 DB 3024 GCGCAGCTCCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3083  
 QY 335 LysValLeuSerLysLeuGlnAlaArgLeuAspAspLeuTyrPheLysPheThrHisSer 354  
 DB 3084 AAGCTTCGACCAAGCTCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3143  
 QY 355 LeuHisAspGlnGlyHisSerHisLeuGlnAspPro 366  
 DB 3144 CTCATCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3179

RESULTS 5  
 US-09-751-877-1  
 Sequence 1, Application US/09751877  
 Patent No. US20020142949A1  
 GENERAL INFORMATION:  
 APPLICANT: Yen, Frances  
 APPLICANT: Dentison, Blake  
 APPLICANT: Bour, Barbara  
 APPLICANT: Bluhm, Bernard  
 APPLICANT: Dumas Milne Edwards, Jean-Baptiste  
 APPLICANT: Duclet, Aymeric  
 APPLICANT: Bouguet-Jeret, Lydie  
 APPLICANT: Ebbets-Keed, Dana  
 APPLICANT: Salter-Cid, Luisa  
 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING NEOPLASTIC CELL GROWTH  
 FILE REFERENCE: 89 US3 REG  
 CURRENT APPLICATION NUMBER: US/09/751,877  
 CURRENT FILING DATE: 2000-12-28  
 NUMBER OF SEQ ID NOS: 6  
 SOFTWARE: PatentL.pm  
 SEQ ID NO:  
 LENGTH: 81001  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: 10946..12946  
 OTHER INFORMATION: 5'regulatory region  
 NAME/KEY: exon  
 LOCATION: 12947..12958  
 OTHER INFORMATION: exon 1  
 NAME/KEY: exon  
 LOCATION: 13470..13526  
 OTHER INFORMATION: exon 2  
 NAME/KEY: exon  
 LOCATION: 13641..13752  
 OTHER INFORMATION: exon 3  
 NAME/KEY: exon  
 LOCATION: 14271..15968  
 OTHER INFORMATION: exon 4  
 NAME/KEY: misc\_feature  
 LOCATION: 15969..17969  
 OTHER INFORMATION: 3'regulatory region  
 NAME/KEY: allele  
 LOCATION: 1239  
 OTHER INFORMATION: 20-828-311 : polymorphic base C or T  
 NAME/KEY: allele  
 LOCATION: 12347  
 OTHER INFORMATION: 17-42-319 : polymorphic base C or T  
 NAME/KEY: allele  
 LOCATION: 15241  
 OTHER INFORMATION: 17-41-250 : polymorphic base C or T  
 NAME/KEY: allele  
 LOCATION: 42218  
 OTHER INFORMATION: 20-841-149 : polymorphic base A or G  
 NAME/KEY: allele  
 LOCATION: 45442



```

OTHER INFORMATION: 20-842-115 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 77058
OTHER INFORMATION: 20-853-415 : polymorphic base C or T
NAME/KEY: primer_bind
LOCATION: 929..949
OTHER INFORMATION: 20-828.pu
NAME/KEY: primer_bind
LOCATION: 1357..1377
OTHER INFORMATION: 20-828..rp complement
NAME/KEY: primer_bind
LOCATION: 12029..12050
OTHER INFORMATION: 17-42.pu
NAME/KEY: primer_bind
LOCATION: 12581..12603
OTHER INFORMATION: 17-42..rp complement
NAME/KEY: primer_bind
LOCATION: 14992..15012
OTHER INFORMATION: 17-41.pu
NAME/KEY: primer_bind
LOCATION: 15460..15482
OTHER INFORMATION: 17-41..rp complement
NAME/KEY: primer_bind
LOCATION: 42070..42090
OTHER INFORMATION: 20-841.pu
NAME/KEY: primer_bind
LOCATION: 42572..42591
OTHER INFORMATION: 20-841..rp complement
NAME/KEY: primer_bind
LOCATION: 45328..45347
OTHER INFORMATION: 20-842.pu
NAME/KEY: primer_bind
LOCATION: 45863..45883
OTHER INFORMATION: 20-842..rp complement
NAME/KEY: primer_bind
LOCATION: 76644..76664
OTHER INFORMATION: 20-853.pu
NAME/KEY: primer_bind
LOCATION: 77166..77185
OTHER INFORMATION: 20-853..rp complement
NAME/KEY: primer_bind
LOCATION: 1220..1238
OTHER INFORMATION: 20-828-311.mis
NAME/KEY: primer_bind
LOCATION: 1240..1258
OTHER INFORMATION: 20-828-311.mis complement
NAME/KEY: primer_bind
LOCATION: 12328..12346
OTHER INFORMATION: 17-42-319.mis
NAME/KEY: primer_bind
LOCATION: 12348..12366
OTHER INFORMATION: 17-42-319.mis complement
NAME/KEY: primer_bind
LOCATION: 15222..15240
OTHER INFORMATION: 17-41-250.mis
NAME/KEY: primer_bind
LOCATION: 15242..15260
OTHER INFORMATION: 17-41-250.mis complement
NAME/KEY: primer_bind
LOCATION: 42199..42217
OTHER INFORMATION: 20-841-149.mis
NAME/KEY: primer_bind
LOCATION: 42219..42237
OTHER INFORMATION: 20-841-149.mis complement
NAME/KEY: primer_bind
LOCATION: 45423..45441
OTHER INFORMATION: 20-842-115.mis
NAME/KEY: primer_bind
LOCATION: 45443..45461
OTHER INFORMATION: 20-842-115.mis complement
NAME/KEY: primer_bind
LOCATION: 77039..77057
OTHER INFORMATION: 20-853-415.mis

```

```

NAME/KEY: primer_bind
LOCATION: 77059..77077
OTHER INFORMATION: 20-853-415.mis complement
NAME/KEY: misc_binding
LOCATION: 1227..1251
OTHER INFORMATION: 20-828-311.probe
NAME/KEY: misc_binding
LOCATION: 12335..12359
OTHER INFORMATION: 17-42-319.probe
NAME/KEY: misc_binding
LOCATION: 15229..15253
OTHER INFORMATION: 17-41-250.probe
NAME/KEY: misc_binding
LOCATION: 42206..42230
OTHER INFORMATION: 20-841-149.probe
NAME/KEY: misc_binding
LOCATION: 45430..45454
OTHER INFORMATION: 20-842-115.probe
NAME/KEY: misc_binding
LOCATION: 77046..77070
OTHER INFORMATION: 20-853-415.probe
US-09-751-877-1

```

```

Alignment Scores:
Pred. No.: 9 21e-296 Length: 81001
Score: 312.00 Matches: 312
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 85.25% Indels: 0
Gaps: 0
US-09-842-364-3 (1-366) x US-09-751-877-1 (1-81001)

```

```

QY 55 ThrLeuLysAspSerLeuGluGlnAspLeuAsnMetAsnLysPheLeuGluLysLeu 74
DB 14272 ACCCTGAAGACAGCCCTTGAGCAAGACCTCAACATATGAACTCCGAAAAGCTG 14331
QY 75 ArgProLeuSerGlySerGluAlaProArgLeuProGlnAspProValGlyMetArg 94
DB 14332 AGCCCTTGAGTGAGAGAGGAGGCTCCGCTCCACAGAGCCGGTGCGATGCGGCG 14391
QY 95 GlnLeuGlnGluGluLeuGluGluGlnValLysAlaArgLeuGlnProTyrMetAlaGln 114
DB 14392 CAGCTGAGAGAGAGAGTGGAGAGGTGAAGGCTCCGCTCCAGCCCTACATGAGAGGCG 14451
QY 115 HisGluLeuValGlyTyrPasnLeuGluGlyLeuArgGlnGlnLeuLysProTyrTrMet 134
DB 14452 CACGAGCTGGTGGGCTGGGAATTTGGAGGGCTTGGCGAGCACTGAAGCCCTACAGATG 14511
QY 135 AspLeuMetGluGlnValAlaLeuArgValGlnGluLeuGlnGlnLeuArgValVal 154
DB 14512 CATCTGATTCAGCAGCAGTGGCCCTCCGCTCCAGCAGCTTCAGCAGCAGTTCGCTG 14571
QY 155 GlyGluAspThrLysAlaGlnLeuLeuGlyGlyValAspGlnAlaTrpAlaLeuLeuGln 174
DB 14572 GGGGAAGACACCAAGGCCCTGCTGGGGGCGCTGGAGACAGGCTTGGCTTGGTGCAG 14631
QY 175 GlyLeuGlnSerArgValAlaHisIsthrGlyArgPheLysGluLeuPheHisProTyr 194
DB 14632 GGAAGTGCAGAGCCGCGGGGACCAACCGCGGCTTCAAGAGGCTTTCACCCATAC 14691
QY 195 AlaGluSerLeuValSerGlyIleGlyArgHisValGlnGluLeuLysArgSerValAla 214
DB 14692 GCCGAGAGCCTGGTGAAGCCGATCGGGCGCCAGCTGCAGAGCTGCACCGCACTGGGCT 14751
QY 215 ProHisAlaProAlaSerProAlaIleArgLeuSerArgCysValGlnValLeuSerArg 234
DB 14752 CCGCAGCGCCCGCCAGCCCGCGCGCTTCAGTGGCTGGAGGCTGCTCCCGGAAAG 14811
QY 235 LeuThrLeuLysAlaLysAlaLeuHisAlaArgIleGlnGlnAsnLeuAspGlnLeuArg 254
DB 14812 CTCACGCTCAAGGCCAAGGCCCTGCAAGCAGCATCAGCAGAACTGTAAGCAGCTGCC 14871

```

```

QY 255 GUGLULeuSerArqAlaIhealacIYthrGlyThrGluGluValaGlyProAspPro 274
    |||||||
Db 14872 GAAGAGCTTAGCAGAGCCTTTGGAGCAGTGGAGTGAAGAGGCGGCGGAGCC 14931
QY 275 GlnMetLeuSerGluGluValaArqGlnArqLeuGlnAlaIhealArqGlnAspThrTyrl 294
    |||||||
Db 14932 CAGATCTCTCCGAGAGGCTGGCCGAGCAGTTCAGGCTTCCGCGAGCAGCAGTCACTG 14991
QY 295 GlnIleAlaIhealArqThrArqAlaIleAspGlnGluThrGluGluValaGlnGlnGln 314
    |||||||
Db 14992 CAGATAGCTGCTTCACTGCTGCGGATGACACAGAGATGAGAGGTCGACAGCAGCTG 15051
QY 315 AlaProProProGlyHisSerAlaIhealArqProGluPheGlnGlnIntrAspSerGly 334
    |||||||
Db 15052 GCGCCAGCTTCACAGCAGGCGCAGTTCGCGCCAGAGTTTCAACAAACAGACAGTGGC 15111
QY 335 LysValLeuSerLysLeuGlnAlaIhealArqLeuAspAspLeuTrpGluAspIleThrHisSer 354
    |||||||
Db 15112 AAGCTTCTGACCAAGCTGCGAGCCCTTCGATGACCTTGGCGAAGACATCACTCAGC 15171
QY 355 LeuHisAspGlnGlnHisSerHisLeuGlyAspPro 366
    |||||||
Db 15172 CTTGATGACCGAGGCGCACAGCCACTGTGGGGAGCC 15207

RESULT 6
US-09-800-729-40
; Sequence 40, Application US/09800729
; Patent No. US20020068319A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 32 Human secreted proteins
; FILE REFERENCE: P2044p1
; CURRENT APPLICATION NUMBER: US/09/800,729
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: PCT/US00/26013
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 60/155,709
; NUMBER OF SEQ ID NOS: 217
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 40
; LENGTH: 1393
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-800-729-40

Alignment Scores:
Pred. No.: 5,76e-265 Length: 1393
Score: 279.00 Matches: 365
Percent Similarity: 99.46% Conservative: 0
Best Local Similarity: 99.46% Mismatches: 1
Query Match: 76.23% Indels: 2
DB: 10 Gaps: 0

US-09-842-364-3 (1-366) x US-09-800-729-40 (1-1393)
QY 1 MetAlaSerMetAlaValaLeuThrTrpAlaLeuAlaLeuSerAlaIhealAspSerAla 20
    |||||||
Db 64 ATGGCAAGCAATGGCTGCGGCTGCTGAGCTGAGCTTCTTTCAGCGTTTCGCC 123
QY 21 ThrGlnAlaIhealArqLysGlyPheTrpAspTyrPheSerGlnThrSerGlyAspLysGlyArq 40
    |||||||
Db 124 ACCAGCAGCAGGAAGGCTCTGGAAGTCTTACAGCAGCAGGCGGAGCAAGAGGAGG 183
QY 41 ValGluGlnIleHisGlnGlnGlnLysMetAlaArqGluProAlaThrLeuLysAspSerLeu 60
    |||||||
Db 184 GTGGAGCAGATCCATCAGCAGGAAGATGCTGCGAGCCCGCCGAAAGACAGCGCTT 243
QY 61 GlnGlnAspLeuAsnAsnMetAlaLysPheLeuGlnLysLeuArqProLeuSerGlySer 80
    |||||||
Db 244 GAGCAAGAGCTCAACATATGAAACAGTTCCTGGAAGAGGTAGGCTTCTGAGTGGAGC 303
QY 81 GluAlaProArqLeuProGlnAspProValaGlyMetArqArqGlnGlnGlnGluLeu 100

```

```

Db 304 GAGGCTCTCCGCTCCACAGACCCGGTGGGATGCGGGGCGAGCTGCGAGAGATTG 363
    |||||||
QY 101 GlnGluValaLysAlaArqLeuGlnProTyrMetAlaGlnAlaIhealGluLeuValaGlyTyr 120
    |||||||
Db 364 GAGGAGGTGAAGCTCGCCCTCCAGCCCTTACATGCGAGAGCCGAGCAGCTGAGTGGCTG 423
QY 121 AsnLeuGlnGlyLeuArqGlnGlnGlnGlnLysProTyrThrMetAspLeuMetGluGlnVal 140
    |||||||
Db 424 AATTTGAGAGGCTTGGCGGCAACTGMAAGCTTACACATGATGTGATGAGAGAGT 483
QY 141 AlaLeuArqValaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 160
    |||||||
Db 484 GCGCTGCGCTGCGAGAGCTGCGAGAGACAGTTCGCGCTGCGTGGGAGAGCAGCAGAGCC 543
QY 161 GlnLeuLeuGlnGlyValaAspGlnAlaIhealArqLeuGlnGlnGlnGlnGlnGlnGln 180
    |||||||
Db 544 CAGTTCCTGGGGGCGGAGAGAGCTTGGCTTCTGCGAGGAGCTTCAGAGCCGCTG 603
QY 181 ValHisHisThrGlyArqPheLysGluLeuPheHisProTyrAlaGluSerLeuValaSer 200
    |||||||
Db 604 GTGCAGCAGACCGGCGCTTCAAGAGCTTTCACCCATAGCGCGAGAGCTTGGTGGAG 663
QY 201 GlyIleGlyArqGlnValaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 220
    |||||||
Db 664 GCGATCGGCGCGCCAGCTGCGAGAGCTTCACCGAGTCTGCGAGCCGCGCGCGCGCAGC 723
QY 221 ProAlaArqLeuSerArqGlyValaGlnValaLeuSerArqLysLeuThrLeuLysAlaLys 240
    |||||||
Db 724 CCGCGCGCGCTCGAGCTGCTGCGAGAGCTTCTCCGAGAGCTCAAGCTCAAGCGCGCAG 783
QY 241 AlaLeuHisAlaArqIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260
    |||||||
Db 784 GCGCTGCGAGCAGCATCCAGAGAGCTTGGAGCAGCTGCGAGAGCTTATC-AGAGC 842
QY 260 AlaLeuGlnGlyThrGlyThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 280
    |||||||
Db 843 CTTTGCAGGAGCTGGAGCTAGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 902
QY 280 uValaArqGlnArqLeuGlnAlaIhealArqGlnAspThrTyrlLeuGlnIleAlaIheal 300
    |||||||
Db 903 GCGCGCGCAGCAGCTTGGAGCTTCCCGCGAGAGCTTCTGAGAGCTTCTGAGAGCTT 962
QY 300 TargAlaIleAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 320
    |||||||
Db 963 TCGCGCATGAGCAGAGAGCTGAGAGGTCGAGCAGCAGCTGAGAGCTTCTGAGAGCT 1022
QY 320 YHisSerAlaIhealArqProGluPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 340
    |||||||
Db 1023 CCACAGTCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1082
QY 340 uGlnAlaArqLeuAspAspLeuTrpGluAspIleThrHisSerLeuHisAspGlnGln 360
    |||||||
Db 1083 CCAAGCGCGCTGCTGATGAGCTTGGAGAGCAGATCACTACAGCTTATGAGCAGGCGCA 1142
QY 360 sSerHisLeuGlyAspPro 366
    |||||||
Db 1143 CAGCAGTCTGGGGAGCC 1161

RESULT 7
US-09-349-015-1
; Sequence 1, Application US/09349015A
; Patent No. US2002015950A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen Anne
; APPLICANT: Volkmuth, Wayne
; APPLICANT: Walker, Michael
; TITLE OF INVENTION: ATHEROSCLEROSIS-ASSOCIATED GENES
; FILE REFERENCE: PB-0013 US
; CURRENT APPLICATION NUMBER: US/09/349,015A
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0

```



```

|||||
Db 124 CACGAGAGCTGTGAGAGAGCTGCGCTCCCTCCACCTACATGGCAGAGCCGACAGAC 183
Qy 117 LeuValGlyTrpAsnLeuGluGlyLeuArgGlnGlnLeuysProGlyThrMetAspLeu 136
Db 184 CTGGTGGCTGCAATTGGAGGCTTCCGCA-CAACTGAACCCCTAACACATGATGATCTG 242
Qy 137 MetGluGlnValAlaLeuArgValGlnGlnLeuGlnGlnLeuArgValValGlyGlu 156
Db 243 ATGGACAGAGTGGCTGCGCGTGCAGAGCTGCGAGAGAGTGGCGCTGGGGGAA 302
Qy 157 AspThrIysAlaGlnLeuLeuGlyGlyValAspGluAlaTrpAlaLeuGlnGlyLeu 176
Db 303 GACACCAAGGCCCGCTTGGCTGGCGCGCTGGAGAGAGCTTGGCGCTTGGCTGCAAGGAGCTG 362
Qy 177 GlnSerArgValValHisThrGlyArgPheGlyLeuPhe 191
Db 363 CACAGCGCGCTGCTGCACACACCGCGCTTCAAGACCTCTTC 407

```

# RESULT 9

```

US-09-800-729-210
; Sequence 210, Application US/09800729
; Patent No. US20020068319A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 32 Human secreted proteins
; FILE REFERENCE: P2044P1
; CURRENT APPLICATION NUMBER: US/09/800,729
; CURRENT FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: PCT/US00/26013
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 60/1155,709
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 217
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 210
; LENGTH: 331
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-800-729-210

```

```

Alignment Scores:
Pred. No.: 6,97e-24 Length: 331
Score: 33.00 Matches: 33
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.02% Indels: 0
DB: 10 Gaps: 0

```

US-09-842-364-3 (1-366) x US-09-800-729-210 (1-331)

```

Qy 334 GLLYValLeuSerIysLeuGlnAlaArgLeuAspLeuTrpGluAspIleThrHis 353
Db 1 CGCAAGCTTCTGACCAACCTGCCAGCCCGCTGCAATGACCTGTGGGACACATCACTAC 60

```

```

Qy 354 SerLeuHisAspGlnGlnIysSerHisIleuGlyAspPro 366
Db 61 AGCCTCATGACACGAGGCGACACGACCATCTGGGGGACCCC 99

```

## RESULT 10

```

US-09-983-965-4674
; Sequence 4674, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26

```

```

; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 4674
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 14-LIB34-048-Q1-E1-D5
US-09-983-965-4674

```

```

Alignment Scores:
Pred. No.: 2.41 Length: 301
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.46% Indels: 0
DB: 10 Gaps: 0

```

US-09-842-364-3 (1-366) x US-09-983-965-4674 (1-301)

```

Qy 8 LeuThrTrpAlaLeuAlaLeuSer 16
Db 31 CTGACCTGGCTCTGCGCTCTCTCTCA 57

```

## RESULT 11

```

US-09-983-965-4360
; Sequence 4360, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 4360
; LENGTH: 397
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 02-LIB34-007-Q1-E1-A9
US-09-983-965-4360

```

```

Alignment Scores:
Pred. No.: 3.11 Length: 397
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.46% Indels: 0
DB: 10 Gaps: 0

```

US-09-842-364-3 (1-366) x US-09-983-965-4360 (1-397)

```

Qy 8 LeuThrTrpAlaLeuAlaLeuSer 16
Db 25 CTGACCTGGCTCTGCGCTCTCTCTCA 51

```

## RESULT 12

```

US-09-960-352-10825
; Sequence 10825, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:

```

APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Nengbing  
APPLICANT: Byatt, John C.  
APPLICANT: Mathialagan, Nagappan  
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
FILE REFERENCE: 16511.006/37-21(10298)C  
CURRENT APPLICATION NUMBER: US/09/960,352  
CURRENT FILING DATE: 2001-09-24  
NUMBER OF SEQ ID NOS: 15112  
SEQ ID NO 10825  
LENGTH: 405  
TYPE: DNA  
ORGANISM: Bos laurus  
OTHER INFORMATION: Clone ID: 46-111B34-061-Q1-E1-D6  
US-09-960-352-10825

Alignment Scores:  
Pred. No.: 3.16 Length: 405  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.46% Indels: 0  
DB: Gaps: 0

US-09-842-364-3 (1-366) x US-09-960-352-10825 (1-405)

OY 95 GlnLeuGlnGluGluLeuGluGluVal 103  
DB 241 CAGCTGCAGAGAGAGCTGAGAGAGGTG 267

RESULT 13  
US-09-960-352-8211  
Sequence 8211, Application US/09960352  
Patent No. US20020137139A1  
GENERAL INFORMATION:  
APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Nengbing  
APPLICANT: Byatt, John C.  
APPLICANT: Mathialagan, Nagappan  
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
FILE REFERENCE: 16511.006/37-21(10298)C  
CURRENT APPLICATION NUMBER: US/09/960,352  
CURRENT FILING DATE: 2001-09-24  
NUMBER OF SEQ ID NOS: 15112  
SEQ ID NO 8211  
LENGTH: 419  
TYPE: DNA  
ORGANISM: Bos laurus  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (386)  
OTHER INFORMATION: unsure at all n locations  
OTHER INFORMATION: Clone ID: 35-LIB34-020-Q1-E2-A4  
US-09-960-352-8211

Alignment Scores:  
Pred. No.: 3.26 Length: 419  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.46% Indels: 0  
DB: Gaps: 0

US-09-842-364-3 (1-366) x US-09-960-352-8211 (1-419)

OY 8 LeuThrTrpAlaLeuAlaLeuLeuSer 16  
DB 31 CTGACCTGGGCTCGCCCTCTCTCA 57

RESULT 14  
US-09-938-842A-2586

Sequence 2586, Application US/09938842A  
Patent No. US20020160378A1  
GENERAL INFORMATION:  
APPLICANT: Harper, Jeff  
APPLICANT: Krebs, Joel  
APPLICANT: Wang, Xun  
APPLICANT: Zhu, Tong  
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
FILE REFERENCE: SRIPI300-3  
CURRENT APPLICATION NUMBER: US/09/938,842A  
CURRENT FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: US 60/227,866  
PRIOR FILING DATE: 2000-08-24  
PRIOR APPLICATION NUMBER: US 60/264,647  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/300,111  
PRIOR FILING DATE: 2001-06-22  
NUMBER OF SEQ ID NOS: 5379  
SEQ ID NO 2586  
LENGTH: 822  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-938-842A-2586

Alignment Scores:  
Pred. No.: 6.06 Length: 822  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.46% Indels: 0  
DB: Gaps: 0

US-09-842-364-3 (1-366) x US-09-938-842A-2586 (1-822)

OY 95 GlnLeuGlnGluGluLeuGluGluVal 103  
DB 533 CAGCTGCAGAGAGAGCTGAGAGAGT 559

RESULT 15  
US-09-815-242-7404/C  
Sequence 7404, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 7404  
 LENGTH: 2850  
 TYPE: DNA  
 ORGANISM: Helicobacter pylori  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (1)...(2850)  
 US-09-815-242-7404

Alignment Scores:  
 Pred. No.: 19  
 Score: 9.00  
 Percent Similarity: 100.00%  
 Best Local Similarity: 100.00%  
 Query Match: 2.46%  
 DB: 10  
 Length: 2850  
 Matches: 9  
 Conservative: 0  
 Mismatches: 0  
 Indels: 0  
 Gaps: 0

US-09-842-364-3 (1-366) x US-09-815-242-7404 (1-2850)

Qy 171 AAlaLeuLeuGlnGlyLeuGlnSerArg 179  
 ||||||||||||||||||  
 Db 1135 GCACCTCTTCACGCGATTCCATCCCGC 1109

Search completed: January 23, 2003, 19:25:28  
 Job time : 89 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: January 23, 2003, 16:10:07 : Search time 62 Seconds  
(without alignments) 786.608 Million cell updates/sec

Title: US-09-842-364-3

Perfect score: 1877  
Sequence: 1 MASMAVYLVWALTLALSAFSA.....LMEDITHSLDGHSHLGD 366

Scoring table: H/OSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002:\*

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1877	100.0	366	21	AA825778 Human secreted pro
2	1877	100.0	366	22	AAE11932 Human CG122 (or C8
3	1877	100.0	366	22	AAU14554 Human novel protei
4	1877	100.0	366	22	AA820103 Human apolipoprote
5	1877	100.0	366	22	AA8175364 Human secreted pro
6	1877	100.0	400	22	AAE11941 Human lipid metabo
7	1877	100.0	400	22	AAU14318 Human novel protei
8	1871	99.7	366	22	AA890666 Human secreted pro
9	1864	99.3	363	21	AA154552 Amino acid sequenc
10	1858	99.0	363	22	AA890573 Human secreted pro

11	1858	99.0	363	23	AB65475 Human albumin fusi
12	1396	74.4	274	23	AAU86153 Human PRO773 polyp
13	1381	73.6	274	23	ABG34037 Human pro peptidase
14	1293	68.9	367	21	AA154561 Amino acid sequenc
15	1293	68.9	367	21	AA154561 Rat rap3 protein s
16	350	18.6	382	22	AA890663 Human secreted pro
17	328	17.5	396	22	AA890664 Human secreted pro
18	324	17.3	396	23	AAU10864 Human apolipoprote
19	322.5	17.2	396	23	AAU10867 Human apolipoprote
20	321	17.1	396	23	AAU10860 Human apolipoprote
21	321	17.1	396	23	AAU10865 Human apolipoprote
22	321	17.1	396	23	AAU10866 Human apolipoprote
23	321	17.1	396	23	AAU10868 Human apolipoprote
24	321	17.1	396	23	AAU10869 Human apolipoprote
25	320	17.0	396	23	AAU10861 Human apolipoprote
26	320	17.0	396	23	AAU10861 Human apolipoprote
27	318	16.9	396	23	AAU10861 Human apolipoprote
28	317	16.9	396	23	AAU10862 Human apolipoprote
29	316	16.8	396	23	AAU10862 Human apolipoprote
30	316	16.8	396	23	AAU10862 Human apolipoprote
31	316	16.8	396	23	AAU10862 Human apolipoprote
32	316	16.8	396	23	AAU10862 Human apolipoprote
33	315	16.8	396	23	AAU10862 Human apolipoprote
34	315	16.8	396	23	AAU10863 Human apolipoprote
35	313	16.7	396	23	AAU10863 Human apolipoprote
36	310	16.5	396	23	AAU10863 Human apolipoprote
37	309	16.5	396	23	AAU10863 Human apolipoprote
38	309	16.5	396	23	AAU10863 Human apolipoprote
39	308.5	16.4	396	23	AAU10863 Human apolipoprote
40	308.5	16.4	396	23	AAU10863 Human apolipoprote
41	300	16.0	396	23	AAU10863 Human apolipoprote
42	300	16.0	396	23	AAU10863 Human apolipoprote
43	298.5	15.9	391	22	AA890665 Human secreted pro
44	298	15.9	391	22	AA890665 Human secreted pro
45	292	15.6	393	14	AA839490 Human apolipoprote

## ALIGNMENTS

RESULT 1	AA825778	Human: secreted protein; forensic procedure; gene therapy;
ID	AA825778	standard; Protein: 366 AA.
XX	AA825778;	
AC	AA825778;	
XX	28-NOV-2000	(first entry)
DT	28-NOV-2000	(first entry)
XX	Human secreted protein SEQ ID #90.	
XX	Human: secreted protein; forensic procedure; gene therapy;	
KW	Chromosome mapping; cancer; autoimmune disease; cardiovascular disorder;	
KW	Cystic fibrosis; hypothyroidism; immunological disorder; amyloidosis;	
KW	Brain disorder; skeletal muscle disorder; eye disorder; obesity;	
KW	mitochondriopathy; diabetes; atherosclerosis; Alzheimer's disease;	
KW	neurodegenerative disorder; graft rejection; dementia; hyperlipidemia;	
XX	septic shock; impotence.	
XX		
OS	Homo sapiens.	
XX		
PN	MO200037491-A2.	
XX		
PD	29-JUN-2000.	
XX		
XX	20-DEC-1999;	99WO-1802058.
PF	20-DEC-1999;	99WO-1802058.
XX		
PR	22-DEC-1998;	98US-0113686.
XX	25-JUN-1999;	99US-0141032.
PA	(GEST ) GENSET.	
XX		
PI	Bougueret L, Dumas J, Duclert A;	
XX		

DR WPI: 2000-442637/38.  
 DR N-PSDB: AAA87740.  
 XX Polynucleotides and polypeptides encoding proteins with signal  
 XX peptides, useful in diagnostic, forensic, gene therapy and chromosome  
 XX mapping procedures -  
 PS Claim 9: Page 263-264; 306pp; English.  
 XX  
 XX This sequence represents a human secreted protein amino acid sequence.  
 CC The invention relates to sequences AA87725-AA87774 which encode human  
 CC secreted proteins AA825763-B25812. The proteins include signal peptides.  
 CC included in the invention are a host cell containing one of the cDNA  
 CC sequences, and a purified antibody capable of binding to one of the  
 CC secreted proteins. Also contained in the invention are methods for  
 CC storing the sequence data on a computer system, and a method for  
 CC identifying features of the cDNA sequences using a computer programme.  
 CC The cDNAs are useful for expressing secreted proteins or fragments to  
 CC obtain antibodies capable of specifically binding to the secreted  
 CC proteins. The cDNAs may also be useful in diagnostic, forensic, gene  
 CC therapy and chromosome mapping procedures and may be used to design  
 CC expression vectors and secretion vectors. The proteins of the invention  
 CC may be used to treat diseases including cancer, autoimmune diseases,  
 CC cardiovascular disorders, cystic fibrosis, hypothyroidism, immunological  
 CC disorders, amyloidosis, brain disorders, skeletal muscle disorders, eye  
 CC disorders, obesity, mitochondrial cytopathies, diabetes, atherosclerosis,  
 CC neurodegenerative disorders, graft rejection, Alzheimer's disease,  
 CC dementia, hyperlipidaemia, septic shock and impotence.  
 CC  
 XX Sequence 366 AA:  
 SQ  
 Query Match 100.0%; Score 1877; DB 21; Length 366;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-155;  
 Matches 366; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MASMAAVLTWALALSAFSAATQARKGFMYFSOTSGDKGRVROIHOKMARREPAATLKDSL 60  
 DB 1 MASMAAVLTWALALSAFSAATQARKGFMYFSOTSGDKGRVROIHOKMARREPAATLKDSL 60  
 QY 61 EDDLNNMNFLEKRLRPLSGSEAPRLPQDPVGMKRLQLELEEVKARLOPYMAEHLVGM 120  
 DB 61 EDDLNNMNFLEKRLRPLSGSEAPRLPQDPVGMKRLQLELEEVKARLOPYMAEHLVGM 120  
 QY 121 NIEGLRQOLKPYTMDLMRQVALRVQELQRLVYGVGDTYAKQLGCVDEAMALLQGLQSRV 180  
 DB 121 NIEGLRQOLKPYTMDLMRQVALRVQELQRLVYGVGDTYAKQLGCVDEAMALLQGLQSRV 180  
 QY 121 NIEGLRQOLKPYTMDLMRQVALRVQELQRLVYGVGDTYAKQLGCVDEAMALLQGLQSRV 180  
 DB 121 NIEGLRQOLKPYTMDLMRQVALRVQELQRLVYGVGDTYAKQLGCVDEAMALLQGLQSRV 180  
 QY 181 VHTGKFKELFPHYAESLVSGLGRHVOHLHRSVAPAPASPARLSRCYQVLSKRLTIKAK 240  
 DB 181 VHTGKFKELFPHYAESLVSGLGRHVOHLHRSVAPAPASPARLSRCYQVLSKRLTIKAK 240  
 QY 241 ALHARLQONLDOLRELSNAFAGTGTGEEVANGPDPOMLSEEVORLOAFRDYTLQIAAFT 300  
 DB 241 ALHARLQONLDOLRELSNAFAGTGTGEEVANGPDPOMLSEEVORLOAFRDYTLQIAAFT 300  
 QY 301 RAIDQTEVEVQOQLAPPPGHSFAFEFOQTDSGKVLISKIOLARDLWEDITHSLDOGH 360  
 DB 301 RAIDQTEVEVQOQLAPPPGHSFAFEFOQTDSGKVLISKIOLARDLWEDITHSLDOGH 360  
 QY 361 SHLGDP 366  
 DB 361 SHLGDP 366  
 DB 361 SHLGDP 366  
 RESULT 2  
 AAEL1922  
 ID AAEL1922 standard; Protein; 366 AA.  
 XX  
 AC AAEL1922;  
 XX  
 DT 18-DEC-2001 (first entry)  
 XX  
 DE Human CG122 (or C868) protein.

XX  
 KW Human: apolipoprotein; lipase; lipoprotein receptor; ALR; angina;  
 KW cardiovascular disease; lipid metabolism; myocardial infarction;  
 KW cerebral ischaemia; arterial thrombosis; thrombolytic; antilipemic;  
 KW coronary artery thrombosis; cerebral artery thrombosis; stroke;  
 KW intracardiac thrombosis; gene therapy; cardiovascular; vasodilator;  
 KW neuroprotectant; cerebroprotective.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..23  
 FT /label= Signal\_peptide  
 FT Protein 24..366  
 FT /note= "Human mature CG122 (or C868) protein"  
 FT Domain 4..251  
 FT /note= "Apolipoprotein A1/A4/E family domain"  
 FT Domain 75..130  
 FT /note= "Apolipoprotein plasma lipid transport domain"  
 FT Domain 92..142  
 FT /note= "Apolipoprotein E precursor domain"  
 PN WO200179446-A2.  
 PD 25-OCT-2001.  
 PR 16-APR-2001; 2001WO-US12529.  
 PR 14-APR-2000; 2000US-197137P.  
 PR 20-JUN-2000; 2000US-0598042.  
 PR 03-AUG-2000; 2000US-0631451.  
 PR 22-SEP-2000; 2000US-0667298.  
 PR 17-NOV-2000; 2000US-0714936.  
 PA (HYSE-) HYSEQ INC.  
 PI Ballinger DG, Loeb D, Montgomery JR, Tang TY, Zhou P, Goodrich R;  
 PI Liu C, Asundi V, Zhao QA, Wehrman T, Drmanac RT, Men F, Qian XB;  
 PI Wang D;  
 DR WPI: 2001-611724/70.  
 DR N-PSDB: AAD19216.  
 XX Nucleic acids encoding human apolipoproteins, lipases, and lipoprotein  
 PT receptor polypeptides, useful for preventing diagnosing and treating  
 PT lipid metabolism disorders, thrombosis and cardiovascular diseases -  
 XX  
 PS Claim 10: Page 148-150; 266pp; English.  
 CC The invention relates to polynucleotides encoding proteins CG122, CG179,  
 CC CG95, CG121, CG162, CG27, CG153 and CG168 which are related to proteins  
 CC involved in lipid metabolism and cardiovascular disease such as human  
 CC apolipoproteins, lipases and lipoprotein receptor proteins. These DNA  
 CC and protein sequences are useful for treating or preventing disorders  
 CC associated with apolipoproteins, lipases and lipoprotein receptor (ALR)  
 CC expression and for treating lipid metabolism, cardiovascular diseases  
 CC and thrombosis. Antibodies against these proteins are useful for  
 CC determining the presence of or predisposition to a disease associated  
 CC with altered levels of these sequences. ALR polypeptides are also  
 CC useful for identifying agents (agonists and antagonists) that bind to  
 CC them and cells expressing ALR proteins are useful for identifying a  
 CC therapeutic agent for use in treatment of a pathology related to  
 CC aberrant expression or physiological interactions of this polypeptide.  
 CC Vectors comprising these DNA and protein sequences are also useful for  
 CC producing ALR proteins. The nucleic acids and polypeptides of the  
 CC invention are also useful for the treatment of occlusive cardiovascular  
 CC diseases, myocardial infarction, cerebral ischaemia, angina, arterial  
 CC thrombosis, coronary artery thrombosis and cerebral artery thrombosis  
 CC or intracardiac thrombosis and stroke. The nucleotides of the invention  
 CC are used in gene therapy. The present sequence is human CG122 (or C868)  
 CC protein.  
 XX  
 SQ Sequence 366 AA:



Query Match	Similarity	100.0%	Score	1877:	DB	22:	Length	366:
Best/Local	Similarity	100.0%	Pred.	No. 3	9e-155:			
Matches	366:	Conservative	0:	Mismatches	0:	Indels	0:	Gaps
								0
QY	1	MASMAAVLTWALALLSAFSATQARKGFMDYFSQTSQDKGRVEQIHQOKMARPATLKDSL	60					
Db	1	MASMAAVLTWALALLSAFSATQARKGFMDYFSQTSQDKGRVEQIHQOKMARPATLKDSL	60					
QY	61	RODLDNMNKFLEKLRPLSGSEAPRLPDDPVGMRRLOLELEEVKARIQPYMAEAHELVCW	120					
Db	61	RODLDNMNKFLEKLRPLSGSEAPRLPDDPVGMRRLOLELEEVKARIQPYMAEAHELVCW	120					
QY	121	NLEGLRQOLKPYTMDLMEQVALRVQELQELRVYGEDTKAQLLGGVDEAMALLQGLQSRV	180					
Db	121	NLEGLRQOLKPYTMDLMEQVALRVQELQELRVYGEDTKAQLLGGVDEAMALLQGLQSRV	180					
QY	181	VHHTGRFKLEFPRYAESLVSIGTRHVOELHRSVAPAPASPAALSRVQVLSKRLTLTKAK	240					
Db	181	VHHTGRFKLEFPRYAESLVSIGTRHVOELHRSVAPAPASPAALSRVQVLSKRLTLTKAK	240					
QY	241	ALHARIQOQMDLRELSRAFAAGTGTBEACGPPQMLSEVRQRLQAFQDQTVLQIAAFT	300					
Db	241	ALHARIQOQMDLRELSRAFAAGTGTBEACGPPQMLSEVRQRLQAFQDQTVLQIAAFT	300					
QY	301	RAIDQTEEVQOQLAPPPGHSFAFEPFOOTDSGKVLQARLDDLMDITSLHDQGH	360					
Db	301	RAIDQTEEVQOQLAPPPGHSFAFEPFOOTDSGKVLQARLDDLMDITSLHDQGH	360					
QY	361	SHLGDP	366					
Db	361	SHLGDP	366					
RESUR" 3								
AAU14554								
AAU14554	standard; Protein: 366 AA.							
AAU14554								
24-OCT-2001	(first entry)							
Human novel	protein #425.							
Human: novel	protein: Antianaemic; osteopathic; anti-inflammatory;							
immunomodulatory; cytoskeletal; neuroprotective; vulnerability; nootropic;								
anticonvulsant; antiallergic; cerebroprotective; antifungal; antiviral;								
antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;								
thrombolytic; immunogen; antibody; gene therapy; neurological disorder;								
Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;								
tissue regeneration; immune disorder.								
Homo sapiens.								
MO20015437-A2.								
02-AUG-2001.								
25-JAN-2001; 2001WO-US02623.								
25-JAN-2000; 2000US-0491404.								
(HYSE-) HYSEQ INC.								
Tang YT, Liu C, Dirmannac RT;								
WP1: 2001-451939/48.								
N-PSDB: AAS22859.								
Isolated polypeptides useful for treating anti-inflammatory diseases,								
nervous system disorders, and for regenerating bone and cartilage -								
Example 4: Page 871-872; 894pp; English.								

Query Match	Best Local Similarity	Score	DB	Length	Matches	Conservative	Mismatches	Indels	Gaps
1 MASMAVAVLFWALALISAFSATQARKGFWMYFSQTSIGDKGRVQIHQOKMAKRPATLKDSI. 60	100.0%	1877	22	366	366	0	0	0	0
1 MASMAVAVLFWALALISAFSATQARKGFWMYFSQTSIGDKGRVQIHQOKMAKRPATLKDSI. 60	100.0%	1877	22	366	366	0	0	0	0
1 MASMAVAVLFWALALISAFSATQARKGFWMYFSQTSIGDKGRVQIHQOKMAKRPATLKDSI. 60	100.0%	1877	22	366	366	0	0	0	0
61 EDDLNMMKFLFKRLPLSSSEAPRLPDPVGMKRLQLEELVEYKARLQPYMAEAHIELVGM 120	100.0%	1877	22	366	366	0	0	0	0
61 EDDLNMMKFLFKRLPLSSSEAPRLPDPVGMKRLQLEELVEYKARLQPYMAEAHIELVGM 120	100.0%	1877	22	366	366	0	0	0	0
121 NIEGLRQOLKPTPTMDLMEQVALRVQELQQLVNGEDTFAQLLGGVDEMAALLQGLQSRV 180	100.0%	1877	22	366	366	0	0	0	0
121 NIEGLRQOLKPTPTMDLMEQVALRVQELQQLVNGEDTFAQLLGGVDEMAALLQGLQSRV 180	100.0%	1877	22	366	366	0	0	0	0
121 NIEGLRQOLKPTPTMDLMEQVALRVQELQQLVNGEDTFAQLLGGVDEMAALLQGLQSRV 180	100.0%	1877	22	366	366	0	0	0	0
181 VHHGTGFKKLFHPVAFSLVSGIGKHVQELHKSVAHPAPASPARLSKCYOVLISKKLTAKK 240	100.0%	1877	22	366	366	0	0	0	0
181 VHHGTGFKKLFHPVAFSLVSGIGKHVQELHKSVAHPAPASPARLSKCYOVLISKKLTAKK 240	100.0%	1877	22	366	366	0	0	0	0
241 AHHARIQOQLDQLREELISAFAGTGEESAGDDPQMLSEVQORLOAFQODYTLQDIAFT 300	100.0%	1877	22	366	366	0	0	0	0
241 AHHARIQOQLDQLREELISAFAGTGEESAGDDPQMLSEVQORLOAFQODYTLQDIAFT 300	100.0%	1877	22	366	366	0	0	0	0
301 RAIDQETEEVQOQLAPPPPGHSAPAFEEQOTDSGKVLSTLQARLDLMDLMDITHSLHDGII 360	100.0%	1877	22	366	366	0	0	0	0
301 RAIDQETEEVQOQLAPPPPGHSAPAFEEQOTDSGKVLSTLQARLDLMDLMDITHSLHDGII 360	100.0%	1877	22	366	366	0	0	0	0
361 SHLQDP 366	100.0%	1877	22	366	366	0	0	0	0
361 SHLQDP 366	100.0%	1877	22	366	366	0	0	0	0
AAAB20103 standard; Protein; 366 AA.	100.0%	1877	22	366	366	0	0	0	0
AAAB20103 standard; Protein; 366 AA.	100.0%	1877	22	366	366	0	0	0	0
AAAB20103 standard; Protein; 366 AA.	100.0%	1877	22	366	366	0	0	0	0
23-APR-2001 (first entry)	100.0%	1877	22	366	366	0	0	0	0
Human apolipoprotein A-IV-related protein (AAARP).	100.0%	1877	22	366	366	0	0	0	0
Apolipoprotein A-IV-related protein; AAARP; human;	100.0%	1877	22	366	366	0	0	0	0

bi-allelic marker: lipid metabolism; liver related disorder;  
 obesity; diabetes; coronary heart disease; diagnosis: therapy.

Homo sapiens.

MO200100803-A2.

04-JAN-2001.

21-JUN-2000; 2000MO-1B01011.

25-JUN-1999; 99US-0141032.

20-DEC-1999; 99MO-1B02058.

21-DEC-1999; 99US-0469099.

(GEST ) GENSET.

Yen F, Denison B, Bour B, Bihain B, Bougueleret L, Duclet A;

Dumas Milne Edwards J;

WPI: 2001-071485/08.

N-PSDB; AAF30035, AAF30036.

Claim 12; Page 240; 260pp; English.

The present sequence is that of the human apolipoprotein A-IV-related protein (AA4RP). AA4RP is differentially expressed in obese mouse models, indicating a role in lipid metabolism related disorders. It appears to be the human homologue of rat regeneration associated protein (RAP3), which is believed to be involved in liver regeneration, and is likely to have a similar function. It also shows 52% similarity to apolipoprotein A-IV, and is likely to have a similar function. The invention also provides AA4RP cDNAs (see AAF30036) and genomic DNA (see AAF30035), bi-allelic markers identified in the AA4RP gene and from genomic regions flanking the gene, and methods for genotyping a nucleic acid containing 1 or more of the bi-allelic markers. Also provided is a method for detecting a statistical correlation between a bi-allelic marker allele and a phenotype and/or between a bi-allelic marker haplotype and a phenotype. Diagnostic methods are provided for determining whether an individual is at risk of developing a lipid metabolism related disorder and/or a liver related disorder, such as obesity, diabetes and coronary heart disease, and whether the individual suffers from such a disorder as a result of a polymorphism in the AA4RP gene. AA4RP polypeptides can also be used in therapy, to screen for agonist and antagonist compounds, and to raise AA4RP-specific antibodies.

Sequence 366 AA;

Query Match 100.0%; Score 1877; DB 22; Length 366;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-155;  
 Matches 366; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MASMAAVITWALALSAFATGARKGFMDYFSQTSQDKGRVQIHQOKMAREPATLTKSL 60  
 1 MASMAAVITWALALSAFATGARKGFMDYFSQTSQDKGRVQIHQOKMAREPATLTKSL 60  
 61 EODLNNMNFLEKLRPLSGSEAPRLPODPVGNRRLOLELEEVKARLOPYMAEAHELVCW 120  
 61 EODLNNMNFLEKLRPLSGSEAPRLPODPVGNRRLOLELEEVKARLOPYMAEAHELVCW 120  
 121 NLEGLRQOLKPYTMDLMEQVALRVOELQRLRVGVEDTKAOLLGVDAMALLQGLQSRV 180  
 121 NLEGLRQOLKPYTMDLMEQVALRVOELQRLRVGVEDTKAOLLGVDAMALLQGLQSRV 180  
 121 NLEGLRQOLKPYTMDLMEQVALRVOELQRLRVGVEDTKAOLLGVDAMALLQGLQSRV 180  
 181 VHTGREFELPHPAESLVSGIRHVOELHRSVAPARSPARLSRCVOVLSRKLTLLAK 240  
 181 VHTGREFELPHPAESLVSGIRHVOELHRSVAPARSPARLSRCVOVLSRKLTLLAK 240

241 ALHARIQOQLDRLRELSRAFACTGTEGAGPPQMLSEFVQRLOAFRODTYLOIAFT 300

241 ALHARIQOQLDRLRELSRAFACTGTEGAGPPQMLSEFVQRLOAFRODTYLOIAFT 300

301 RAIDETEVEVQOQLAPPFGHSAPAFEPQOTDSGKYSKLQARLDLMDITSLIHQGH 360

301 RAIDETEVEVQOQLAPPFGHSAPAFEPQOTDSGKYSKLQARLDLMDITSLIHQGH 360

361 SHLQDP 366

361 SHLQDP 366

RESULT 5  
 AAB75364  
 ID AAB75364 standard; protein; 366 AA.

AAB75364;

05-APR-2001 (first entry)

Human secreted protein #23.

Secreted protein; prevention; treatment; diagnosis; disease; infection.

Homo sapiens.

MO200100806-A2.

04-JAN-2001.

21-JUN-2000; 2000MO-1B00951.

25-JUN-1999; 99US-0141032.

21-DEC-1999; 99US-0469099.

(GEST ) GENSET.

Dumas Milne Edwards J, Bougueleret L, Jobert S;

WPI: 2001-071487/08.

49 Secreted proteins and the nucleic acids encoding them, useful in gene therapy and for detecting similar sequences in samples -

Claim 10; Page 288-289; 307pp; English.

The present invention relates to 49 secreted proteins and the cDNAs encoding them. The protein and nucleic acids may be used in the prevention, treatment and diagnosis of diseases associated with inappropriate protein expression.

Sequence 366 AA;

Query Match 100.0%; Score 1877; DB 22; Length 366;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-155;  
 Matches 366; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MASMAAVITWALALSAFATGARKGFMDYFSQTSQDKGRVQIHQOKMAREPATLTKSL 60  
 1 MASMAAVITWALALSAFATGARKGFMDYFSQTSQDKGRVQIHQOKMAREPATLTKSL 60  
 61 EODLNNMNFLEKLRPLSGSEAPRLPODPVGNRRLOLELEEVKARLOPYMAEAHELVCW 120  
 61 EODLNNMNFLEKLRPLSGSEAPRLPODPVGNRRLOLELEEVKARLOPYMAEAHELVCW 120  
 121 NLEGLRQOLKPYTMDLMEQVALRVOELQRLRVGVEDTKAOLLGVDAMALLQGLQSRV 180  
 121 NLEGLRQOLKPYTMDLMEQVALRVOELQRLRVGVEDTKAOLLGVDAMALLQGLQSRV 180  
 121 NLEGLRQOLKPYTMDLMEQVALRVOELQRLRVGVEDTKAOLLGVDAMALLQGLQSRV 180  
 181 VHTGREFELPHPAESLVSGIRHVOELHRSVAPARSPARLSRCVOVLSRKLTLLAK 240  
 181 VHTGREFELPHPAESLVSGIRHVOELHRSVAPARSPARLSRCVOVLSRKLTLLAK 240

DB 181 VHHGGRFKELEPHVAESLVSGIGRHVOELHRSVAPHAAPASPARLSRCQVLSRKLTLLAK 240  
 QY 241 ALHARIQONLDOLRELSRAFGCTGTCAGAPDPOMISEEVQRORLOARODTYLQIAAFT 300  
 DB 241 ALHARIQONLDOLRELSRAFGCTGTCAGAPDPOMISEEVQRORLOARODTYLQIAAFT 300  
 QY 301 RAIDQETEEVVOQLAPPPGHSAPAFEPFOOTDSGKVLKLOARLDLMDITTHSLHDGCH 360  
 DB 301 RAIDQETEEVVOQLAPPPGHSAPAFEPFOOTDSGKVLKLOARLDLMDITTHSLHDGCH 360  
 QY 361 SHLGD 366  
 DB 361 SHLGD 366  
 RESULT 6  
 ID AE11941  
 AC AE11941: standard; Protein: 400 AA.  
 DE 18-DEC-2001 (first entry)  
 XX  
 XX Human lipid metabolism related protein #4.  
 XX  
 KM Human; apolipoprotein; lipase; lipoprotein receptor; ALR; angina;  
 KM cardiovascular disease; lipid metabolism; myocardial infarction;  
 KM cerebral ischaemia; arterial thrombosis; thrombolytic; antilipemic;  
 KM coronary artery thrombosis; cerebral artery thrombosis; stroke;  
 KM intracardiac thrombosis; gene therapy; cardiovascular; vasodilator;  
 KM neuroprotectant; cerebroprotective.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200179446-A2.  
 XX  
 PD 25-OCT-2001.  
 XX  
 PF 16-APR-2001; 2001WO-US12529.  
 XX  
 PR 14-APR-2000; 2000US-197137P.  
 PR 20-JUN-2000; 2000US-059804Z.  
 PR 03-AUG-2000; 2000US-0631451.  
 PR 22-SEP-2000; 2000US-0667298.  
 PR 17-NOV-2000; 2000US-0714936.  
 XX  
 PA (HXSE-) HXSEQ INC.  
 XX  
 PI Ballinger DG, Loeb D, Montgomery JR, Tang TY, Zhou P, Goodrich R;  
 PI Liu C, Asundi V, Zhao QA, Wehrman T, Drmanac RT, Ren F, Qian XB;  
 PI Wang D;  
 DR WPI: 2001-611724/70.  
 DR N-PSDB: AAD19236.  
 XX  
 PT Nucleic acids encoding human apolipoproteins, lipases, and lipoprotein  
 PT receptor polypeptides, useful for preventing diagnosing and treating  
 PT lipid metabolism disorders, thrombosis and cardiovascular diseases -  
 XX  
 PS Claim 10; Page 260-261; 266pp; English.  
 XX  
 CC The invention relates to polynucleotides encoding proteins CG122, CG179,  
 CC CG95, CG121, CG162, CG27, CG153 and CG168 which are related to proteins  
 CC involved in lipid metabolism and cardiovascular disease such as human  
 CC apolipoproteins, lipases and lipoprotein receptor proteins. These DNA  
 CC and protein sequences are useful for treating or preventing disorders  
 CC associated with apolipoproteins, lipases and lipoprotein receptor (ALR)  
 CC expression and for treating lipid metabolism, cardiovascular diseases  
 CC and thrombosis. Antibodies against these proteins are useful for  
 CC determining the presence of or predisposition to a disease associated  
 CC with altered levels of these sequences. ALR polypeptides are also  
 CC useful for identifying agents (agonists and antagonists) that bind to

CC them and cells expressing ALR proteins are useful for identifying a  
 CC therapeutic agent for use in treatment of a pathology related to  
 CC aberrant expression or physiological interactions of this polypeptide.  
 CC Vectors comprising these DNA and protein sequences are also useful for  
 CC producing ALR proteins. The nucleic acids and polypeptides of the  
 CC invention are also useful for the treatment of occlusive cardiovascular  
 CC diseases, myocardial infarction, cerebral ischemia, angina, arterial  
 CC thrombosis, coronary artery thrombosis and cerebral artery thrombosis  
 CC or intracardiac thrombosis and stroke. The nucleotides of the invention  
 CC are used in gene therapy. The present sequence is human protein  
 CC related to proteins involved in lipid metabolism.  
 CC  
 XX  
 SQ Sequence 400 AA:  
 Query Match 100.0%; Score 1877; DB 22; Length 400;  
 Matches 366; Conservativity 100.0%; Pred. No. 4.4e-155;  
 Matches 366; Conservativity 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MASMAAVLTWALALLSAFSAATQARKGFMDYFSQTSQDKGRVEQIHQKMAREPATLKDSL 60  
 DB 35 MASMAAVLTWALALLSAFSAATQARKGFMDYFSQTSQDKGRVEQIHQKMAREPATLKDSL 94  
 QY 61 EODLNNMKNKFLKLRPLSGSEAPRLPDPVCKRROLQELIEVKARLOPYMAEAHELVGW 120  
 DB 95 EODLNNMKNKFLKLRPLSGSEAPRLPDPVCKRROLQELIEVKARLOPYMAEAHELVGW 154  
 QY 121 NIEGIRROQLKPYTMDLMOVALRVOELRVEGEDTKAOLIGVDHAMAALQGLOSRV 180  
 DB 155 NIEGIRROQLKPYTMDLMOVALRVOELRVEGEDTKAOLIGVDHAMAALQGLOSRV 214  
 QY 181 VHHGGRFKELEPHVAESLVSGIGRHVOELHRSVAPHAAPASPARLSRCQVLSRKLTLLAK 240  
 DB 215 VHHGGRFKELEPHVAESLVSGIGRHVOELHRSVAPHAAPASPARLSRCQVLSRKLTLLAK 274  
 QY 241 ALHARIQONLDOLRELSRAFGCTGTCAGAPDPOMISEEVQRORLOARODTYLQIAAFT 300  
 DB 275 ALHARIQONLDOLRELSRAFGCTGTCAGAPDPOMISEEVQRORLOARODTYLQIAAFT 334  
 QY 301 RAIDQETEEVVOQLAPPPGHSAPAFEPFOOTDSGKVLKLOARLDLMDITTHSLHDGCH 360  
 DB 335 RAIDQETEEVVOQLAPPPGHSAPAFEPFOOTDSGKVLKLOARLDLMDITTHSLHDGCH 394  
 QY 361 SHLGD 366  
 DB 395 SHLGD 400  
 XX  
 RESULT 7  
 ID AAU14318  
 AC AAU14318: standard; Protein: 400 AA.  
 DE 24-OCT-2001 (first entry)  
 XX  
 DE Human novel protein #189.  
 XX  
 KM Human; novel protein; antihaemic; osteopathic; antiinflammatory;  
 KM immunomodulatory; cytosolic; neuroprotective; vulnary; nootropic;  
 KM anticonvulsant; antiallergic; cerebroprotective; antitumor; antiviral;  
 KM antibacterial; antiallergic; dermatological; haemostatic; antisthmatic;  
 KM thrombolytic; immunogen; antibody; gene therapy; neurological disorder;  
 KM Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;  
 KM tissue regeneration; immune disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200155437-A2.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PF 25-JAN-2001; 2001WO-US02623.  
 XX



Db 61 EDDLNNMNFLEKLRPLSGSEAPRLPQDPVGMRRQLQEELEEVKARLQPYMAEAHELVCW 120  
 QY 121 NIEGLRQOLKPYTMDLMEQVALRVQELQEDTKAQLLGGVDEAMALLQGLQSRV 180  
 Db 121 NIEGLRQOLKPYTMDLMEQVALRVQELQEDTKAQLLGGVDEAMALLQGLQSRV 180  
 QY 181 VHTGRKELFHPYAESLVSIGIRHVOELHRSVAAPHAPASPARLSRCVYLKRLTKAK 240  
 Db 181 VHTGRKELFHPYAESLVSIGIRHVOELHRSVAAPHAPASPARLSRCVYLKRLTKAK 240  
 QY 241 ALHARIQONLDOLREELSRPAFTGTEGAGPPOMLSEEVROLDARFODTYLQIAFT 300  
 Db 241 ALHARIQONLDOLREELSRPAFTGTEGAGPPOMLSEEVROLDARFODTYLQIAFT 300  
 QY 301 RAIDQEEVQOOLAPPFGHSAFAPEFOOTDSGKVLKLOARLDLMDITISLHDQGH 360  
 Db 301 RAIDQEEVQOOLAPPFGHSAFAPEFOOTDSGKVLKLOARLDLMDITISLHDQGH 360  
 QY 361 SHLGDP 366  
 Db 361 SHLGDP 366

RESULT 9  
 ID AAY54562 standard; protein: 363 AA.  
 AC AAY54562;  
 DT 25-APR-2000 (first entry)

DE Amino acid sequence of the human RAP3 protein.  
 KW RAP3; regeneration association protein 3; liver regeneration;  
 KM liver proliferation.  
 OS Homo sapiens.

PN WO200003013-A2.  
 PD 20-JAN-2000.  
 PP 12-JUL-1999; 99WO-EP04938.  
 PR 10-JUL-1998; 98EP-0202336.

PI (AMST-) AMSTERDAM MOLECULAR THERAPEUTICS BV.  
 PI Chamuleau RAFM, Groenink M, Van Der Vijet HN, Iecgwater ACJ;  
 DR WPI: 2000-147615/13.  
 DR N-PSDB: AA245874, AA245875.

XX Isolated RAP3 gene, protein and antibody useful for diagnosing liver  
 PT regeneration and/or cell proliferation -  
 PS Claim 6; Fig 7; 42pp; English.

CC The present sequence represents a human RAP3 (regeneration association  
 CC protein 3). The RAP3 cDNA sequence was isolated from a human liver cDNA  
 CC library. The RAP3 gene is involved in regeneration processes of the  
 CC liver. The rat RAP3 gene was found to be upregulated 6 hours after  
 CC partial hepatectomy, after which it was downregulated. The RAP3 cDNA  
 CC sequence is useful as a source for PCR primers and probes, which are  
 CC useful for detecting nucleotide sequences in a source material. The  
 CC RAP3 cDNA sequence is useful as a marker of liver proliferation. The  
 CC RAP3 protein is useful for the diagnosis of liver regeneration and liver  
 CC cell proliferation. RAP3 antibodies, PCR primers and probes are useful  
 CC for detecting the occurrence of liver cell proliferation in a patient.  
 CC The RAP3 protein is also useful for enhancing the growth of  
 CC regeneration of liver tissue comprising treating the liver tissue  
 CC such as extracorporeal or intracorporeal.

SQ Sequence 363 AA;

Query Match 99.3%; Score 1864; DB 21; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 5.3e-154;  
 Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 MAAVLTWALALLSFASTQARKGFMDYFSQTSQDKGVEDIHOOKKAREPATLKDSLEOD 63  
 Db 1 MAAVLTWALALLSFASTQARKGFMDYFSQTSQDKGVEDIHOOKKAREPATLKDSLEOD 60  
 QY 64 LNNMNFLEKLRPLSGSEAPRLPQDPVGMRRQLQEELEEVKARLQPYMAEAHELVCW 123  
 Db 61 LNNMNFLEKLRPLSGSEAPRLPQDPVGMRRQLQEELEEVKARLQPYMAEAHELVCW 120  
 QY 124 GLRQOLKPYTMDLMEQVALRVQELQEDTKAQLLGGVDEAMALLQGLQSRV 183  
 Db 121 GLRQOLKPYTMDLMEQVALRVQELQEDTKAQLLGGVDEAMALLQGLQSRV 180  
 QY 184 TGRKELFHPYAESLVSIGIRHVOELHRSVAAPHAPASPARLSRCVYLKRLTKAK 243  
 Db 181 TGRKELFHPYAESLVSIGIRHVOELHRSVAAPHAPASPARLSRCVYLKRLTKAK 240  
 QY 244 ARIQONLDOLREELSRPAFTGTEGAGPPOMLSEEVROLDARFODTYLQIAFT 303  
 Db 241 ARIQONLDOLREELSRPAFTGTEGAGPPOMLSEEVROLDARFODTYLQIAFT 300  
 QY 304 DQTEEEVQOOLAPPFGHSAFAPEFOOTDSGKVLKLOARLDLMDITISLHDQGH 363  
 Db 301 DQTEEEVQOOLAPPFGHSAFAPEFOOTDSGKVLKLOARLDLMDITISLHDQGH 360  
 QY 364 GDP 366  
 Db 361 GDP 363

RESULT 10  
 ID AAB90573 standard; protein: 363 AA.  
 AC AAB90573;  
 DT 01-JUN-2001 (first entry)

DE Human secreted protein, SEQ ID NO: 111.

KW Human; secreted protein; immunomodulatory; antileukemic;  
 KW dermatological; anti-inflammatory; anti-HIV; cytostatic; cardiant;  
 KW vascular; anti-angiogenic; ophthalmological; neuroprotectant;  
 KW neurotropic; anticonvulsant; anti-alzheimers; antiparkinsonian;  
 KW antimicrobial; vulnecary; vaccine; gene therapy; cancer;  
 KW protein coordinate data; infection.

XX Homo sapiens.  
 OS  
 PN WO200121658-A1.  
 PD 29-MAR-2001.

PF 22-SEP-2000; 2000WO-US26013.  
 XX  
 PP 24-SEP-1999; 99US-0155709.

PA (HUMA-) HUMAN GENOME SCI INC.

PI N.J. Baker KP, Birge CE, Ebner R, Fiscella M, Komatsoulis GA;  
 PI Lafleur DW, Moore PA, Olsen HS, Rosen CA, Ruben SA, Soppet DR;  
 PI Young PE, Wei P, Florence KA;

DR WPI: 2001-235311/24.  
 DR N-PSDB: AAF97913.

PT Nucleic acids encoding 32 human secreted polypeptides, useful for  
 PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's

PT disease and diabetic retinopathy -  
 XX  
 PS Claim 11; Page 794-795; 890pp; English.  
 CC The present sequence is one of 32 novel human secreted polypeptides. The  
 CC nucleic acid molecules and polypeptides may be used in the prevention,  
 CC diagnosis and treatment of diseases such as immune disorders (e.g.  
 CC multiple sclerosis, systemic lupus erythematosus and human  
 CC immunodeficiency virus (HIV) infections), hyperproliferative disorders  
 CC (e.g. cancers and Gaucher's disease), cardiovascular diseases  
 CC (e.g. Schlimmer syndrome, Chaga's cardiomyopathy and coronary  
 CC arteriosclerosis), angiogenic disorders (e.g. corneal graft  
 CC neovascularisation and diabetic retinopathy), neurological disorders  
 CC (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),  
 CC infectious diseases and/or for promoting wound healing, regeneration  
 CC and/or chemotaxis. The nucleic acid molecules may be used to produce the  
 CC secreted polypeptides. They may also be used as DNA probes in diagnostic  
 CC assays to detect and quantitate the presence of similar nucleic acid  
 CC sequences in samples. The polypeptides may be used as antigens in the  
 CC production of antibodies and in assays to identify modulators of  
 CC their expression and activity.  
 CC  
 SQ Sequence 363 AA:

Query Match 99.0%; Score 1858; DB 22; Length 363;  
 Best Local Similarity 99.7%; Pred. No. 1,8e-153;  
 Matches 362; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 MAAVLTWALALISAFSATOARKGFMDYFSQTSQDKGRVEQIHQOKKAREPATLKDSLEOD 63  
 DB 1 MAAVLTWALALISAFSATOARKGFMDYFSQTSQDKGRVEQIHQOKKAREPATLKDSLEOD 60  
 QY 64 LNNMNFLEKLRPLSGSEAPRLPQDPVGMRRQLQELFEVKARLQPYMAFAHELVGMNLE 123  
 DB 61 LNNMNFLEKLRPLSGSEAPRLPQDPVGMRRQLQELFEVKARLQPYMAFAHELVGMNLE 120  
 QY 124 GLRQOLKPYTMDLMEQVALRVQELQOLRVYCGEDTKAQLLGVDKMAALLQGLQSRVYHH 183  
 DB 121 GLRQOLKPYTMDLMEQVALRVQELQOLRVYCGEDTKAQLLGVDKMAALLQGLQSRVYHH 180  
 QY 184 TGRFKELFHPYAESLVSIGIRHVOELHRSVAPAPASPARLSKCYOVSRLKTLKAKAH 243  
 DB 181 TGRFKELFHPYAESLVSIGIRHVOELHRSVAPAPASPARLSKCYOVSRLKTLKAKAH 240  
 QY 244 ARIQONLQDLREBLRAFAFGTGBGAGDPQMLSEVQRQLQAFRODTYLIQIAAFTTAAI 303  
 DB 241 ARIQONLQDLREBLRAFAFGTGBGAGDPQMLSEVQRQLQAFRODTYLIQIAAFTTAAI 300  
 QY 304 DQETEEVOOQLAPPPGGHSAFAPEFOQTDGKVLKQLARLDLMDLWEDITHSLHDOGHSHL 363  
 DB 301 DQETEEVOOQLAPPPGGHSAFAPEFOQTDGKVLKQLARLDLMDLWEDITHSLHDOGHSHL 360  
 QY 364 GDP 366  
 DB 361 GDP 363

RESULT 11  
 ABG65475  
 ID ABG65475 standard; Protein: 363 AA.

XX ABG65475;  
 DT 27-AUG-2002 (first entry)

XX Human albumin fusion protein #2150.

XX Albumin fusion protein; therapeutic protein X; human albumin; HA;  
 KW human serum albumin; HSA; cancer; reproductive disorder;  
 KW digestive disorder; immune disorder; endocrine disorder;  
 KW haematopoietic disorder; neural disorder; connective disorder;  
 KW cytostatic; antifertility; antiinflammatory; anticancer;  
 KW immunomodulator; anti-HIV; antidiabetic; haemostatic; nootropic;

KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;  
 KW osteopathic; antiarthritic.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 PN WO20017137-A1.  
 XX  
 XX 18-OCT-2001.  
 PD  
 XX 12-APR-2001; 2001WO-US11988.  
 XX  
 PR 12-APR-2000; 2000US-229358P.  
 PR 25-APR-2000; 2000US-199384P.  
 PR 21-DEC-2000; 2000US-256931P.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 PI Rosen CA, Haselaine WA;  
 PI  
 PI WPI; 2002-010886/01.  
 DK  
 XX  
 XX  
 PT New fusion protein for treating disease e.g. diabetes comprises an  
 PT albumin fused to a therapeutic protein -  
 XX  
 PS Claim 1; Page 2047-2049; 2102pp; English.

The present invention relates to albumin fusion proteins comprising a  
 CC therapeutic protein X and human albumin (HA, also known as human serum  
 CC albumin, HSA). The proteins are useful for treating a disease or  
 CC disorder that may be modulated by therapeutic protein X. The albumin  
 CC extends the shelf-life of protein X, and may increase its biological  
 CC in vitro/in vivo activity. The protein is useful for treating and  
 CC diagnosing disorders such as cancer, reproductive disorders, digestive  
 CC disorders (e.g. Crohn's disease, ulcerative colitis), immune disorders  
 CC (e.g. acquired immunodeficiency syndrome, AIDS), endocrine disorders  
 CC (e.g. diabetes), haematopoietic disorders, neural disorders  
 CC (e.g. Alzheimer's, Parkinson's, Creutzfeldt-Jacob disease,  
 CC encephalomyelitis, meningitis, schizophrenia), and connective disorders  
 CC (e.g. osteoporosis, arthritis). ABG63326-ABG65318 represent albumin  
 CC fusion proteins of the invention.  
 CC  
 SQ Sequence 363 AA:

Query Match 99.0%; Score 1858; DB 23; Length 363;  
 Best Local Similarity 99.7%; Pred. No. 1,8e-153;  
 Matches 362; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 MAAVLTWALALISAFSATOARKGFMDYFSQTSQDKGRVEQIHQOKKAREPATLKDSLEOD 63  
 DB 1 MAAVLTWALALISAFSATOARKGFMDYFSQTSQDKGRVEQIHQOKKAREPATLKDSLEOD 60  
 QY 64 LNNMNFLEKLRPLSGSEAPRLPQDPVGMRRQLQELFEVKARLQPYMAFAHELVGMNLE 123  
 DB 61 LNNMNFLEKLRPLSGSEAPRLPQDPVGMRRQLQELFEVKARLQPYMAFAHELVGMNLE 120  
 QY 124 GLRQOLKPYTMDLMEQVALRVQELQOLRVYCGEDTKAQLLGVDKMAALLQGLQSRVYHH 183  
 DB 121 GLRQOLKPYTMDLMEQVALRVQELQOLRVYCGEDTKAQLLGVDKMAALLQGLQSRVYHH 180  
 QY 184 TGRFKELFHPYAESLVSIGIRHVOELHRSVAPAPASPARLSKCYOVSRLKTLKAKAH 243  
 DB 181 TGRFKELFHPYAESLVSIGIRHVOELHRSVAPAPASPARLSKCYOVSRLKTLKAKAH 240  
 QY 244 ARIQONLQDLREBLRAFAFGTGBGAGDPQMLSEVQRQLQAFRODTYLIQIAAFTTAAI 303  
 DB 241 ARIQONLQDLREBLRAFAFGTGBGAGDPQMLSEVQRQLQAFRODTYLIQIAAFTTAAI 300  
 QY 304 DQETEEVOOQLAPPPGGHSAFAPEFOQTDGKVLKQLARLDLMDLWEDITHSLHDOGHSHL 363  
 DB 301 DQETEEVOOQLAPPPGGHSAFAPEFOQTDGKVLKQLARLDLMDLWEDITHSLHDOGHSHL 360  
 QY 364 GDP 366

DB 361 GDP 363

RESULT 12  
AAU86153  
ID AAU86153 standard; Protein; 274 AA.

XX AAU86153;

DT 15-JUL-2002 (first entry)

DE Human PRO773 polypeptide.

KW Human; PRO; benign tumour; malignant tumour; lymphoid malignancy;

KW Leukemia; neuronal disorder; stromal disorder; blastocellular disorder;

KW Inflammatory disorder; immune disorder; angiogenic disorder;

KW Cystostatic; neuroprotective.

OS Homo sapiens.

PN WO200153486-A1.

PD 26-JUL-2001.

PF 11-FEB-2000; 2000WO-US03565.

PR 08-MAR-1999; 99WO-US05028.

PR 11-MAR-1999; 99US-123972P.

PR 11-MAY-1999; 99US-133459P.

PR 02-JUN-1999; 99WO-US12252.

PR 22-JUN-1999; 99US-140650P.

PR 20-JUL-1999; 99US-144758P.

PR 26-JUL-1999; 99US-145688P.

PR 28-JUL-1999; 99US-146222P.

PR 17-AUG-1999; 99US-149395P.

PR 31-AUG-1999; 99US-151689P.

PR 01-SEP-1999; 99WO-US20111.

PR 15-SEP-1999; 99WO-US21090.

PR 30-NOV-1999; 99WO-US28313.

PR 01-DEC-1999; 99WO-US28301.

PR 01-DEC-1999; 99WO-US28634.

PR 05-JAN-2000; 2000WO-US00219.

XX (GETH ) GENENTECH INC.

XX Ashkenazi AJ, Goddard A, Godowski PJ, Gurney AL, Hillan KJ;

PI Marsters SA, Pan J, Pitti RM, Roy MA, Smith V, Stone DM;

PI Watanabe CK, Wood WI;

DR WPI: 2002-205567/26.

DR N-PSDB; ABK40279.

PT Thirty five nucleic acids encoding PRO polypeptides, useful for

PT treating benign or malignant tumours, leukaemias and lymphoid

PT malignancies, inflammatory, angiogenic and immunologic disorders -

PS Claim 61; Fig 52; 302pp; English.

XX The present invention relates to the isolation of novel human PRO

CC polypeptides and the polynucleotide sequences encoding them. The

CC PRO polypeptides, agonists, antagonists or anti-PRO antibodies are

CC useful for treating benign or malignant tumours (e.g. renal, kidney,

CC bladder, breast, etc), leukaemias and lymphoid malignancies, other

CC disorders such as neuronal, glial, astrocytal, hypothalamic, glandular,

CC macrophagal, stromal and blastocellular disorders, inflammatory, immune

CC and angiogenic disorders. The polynucleotide sequences are also

CC useful in gene therapy. AAU86128-AAU86162 represent the human PRO

CC polypeptides of the invention.

XX Sequence 274 AA;

Query Match 74.4%; Score 1396; DB 23; Length 274;  
Best Local Similarity 100.0%; Pred. No. 2; 1e-113;  
Matches 274; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASHAAVLTWALALLSFASTQARKGFMDYFSQTSQDKGRVEQIHQOKMAREPATLKDLSL 60

DB 1 MASHAAVLTWALALLSFASTQARKGFMDYFSQTSQDKGRVEQIHQOKMAREPATLKDLSL 60

QY 61 EODLNNNNKFLKRLPLSGSEAPRLPDDPYGMRQLODELEEVNARLQPTMAEHELGVG 120

DB 61 EODLNNNNKFLKRLPLSGSEAPRLPDDPYGMRQLODELEEVNARLQPTMAEHELGVG 120

QY 121 NLEGLRQQLKPYTMDLMEQVALRVQLOEQLRVYGEDTKAQLGCVDEANALJQGLQSRV 180

DB 121 NLEGLRQQLKPYTMDLMEQVALRVQLOEQLRVYGEDTKAQLGCVDEANALJQGLQSRV 180

QY 181 VHHTRGKFLFHPYAESLVSGIGHVQELHRSVAPHA PASPARLSKCVYLTKAK 240

DB 181 VHHTRGKFLFHPYAESLVSGIGHVQELHRSVAPHA PASPARLSKCVYLTKAK 240

QY 241 ALHARIQNDQLREELSRFAFGTTEGAGPDP 274

DB 241 ALHARIQNDQLREELSRFAFGTTEGAGPDP 274

RESULT 13

ABG34037

ID ABG34037 standard; Protein; 274 AA.

AC ABG34037;

DT 15-JUL-2002 (first entry)

DE Human Pro peptide #8.

KW Human; PRO; secreted protein; transmembrane protein;

KW genetic disorder; tumour; cancer.

OS Homo sapiens.

PN WO200224888-A2.

PD 28-MAR-2002.

PF 29-AUG-2001; 2001WO-US27099.

PR 01-SEP-2000; 2000US-229896P.

PR 05-SEP-2000; 2000US-230621P.

PR 22-SEP-2000; 2000US-235147P.

PR 10-NOV-2000; 2000WO-US10873.

PR 12-JAN-2001; 2001US-261878P.

PR 16-JAN-2001; 2001US-261910P.

PR 16-JAN-2001; 2001US-261939P.

PR 16-JAN-2001; 2001US-262150P.

PR 25-JAN-2001; 2001US-264395P.

PR 02-FEB-2001; 2001US-266421P.

PR 09-FEB-2001; 2001US-267623P.

PR 28-FEB-2001; 2001WO-US05520.

PR 09-MAR-2001; 2001US-274399P.

PR 03-APR-2001; 2001US-280982P.

PR 04-APR-2001; 2001US-282129P.

PR 09-MAY-2001; 2001US-290589P.

PR 25-MAY-2001; 2001WO-US17092.

PR 01-JUN-2001; 2001WO-US17800.

PR 20-JUN-2001; 2001WO-US19692.

PR 29-JUN-2001; 2001WO-US21066.

PR 09-JUL-2001; 2001WO-US21735.

XX (GETH ) GENENTECH INC.

XX Baker KP, Eaton DL, Filvaroff E, Goddard A, Grimaldi JC;

PI Gurney AL, Smith V, Stephan J, Watanabe CK, Wood WI, Zhang Z;



PI Fong S;  
 XX WPI: 2002-362426/39.  
 DR N-PSDB; ABR69968.  
 XX  
 XX New PRO polypeptides and polynucleotides encoding the polypeptides,  
 PT useful in gene therapy, chromosome identification, tissue typing, or  
 PT for genetic analysis of individuals with genetic disorders -  
 XX  
 PS Claim 11; Figure 16; 218pp: English.

XX This invention relates to the cDNA and protein sequences of novel  
 CC secreted and transmembrane polypeptides PRO polypeptides. The  
 CC invention also comprises a method for producing the proteins of the  
 CC invention by recombinant means and antibodies specific for the protein  
 CC of the invention. The antibody may be used for detecting the PRO  
 CC proteins of the invention and may be used to modify their activity.  
 CC polynucleotides may be used as hybridisation probes for a cDNA library  
 CC to isolate the full-length PRO cDNA or to isolate other cDNAs, to  
 CC construct hybridisation probes for mapping the gene which encodes that  
 CC PRO and for genetic analysis of individuals with genetic disorders, in  
 CC assays to identify other proteins or molecules involved in binding  
 CC reaction, to generate transgenic animals or knock-out animals which in  
 CC turn are useful in the development and screening of therapeutically  
 CC useful reagents, for chromosome identification, and tissue typing. The  
 CC PRO polypeptides are useful in gene therapy, and as molecular weight  
 CC markers for protein electrophoresis purposes. The sequences may  
 CC also be used to detect overexpression on PRO polypeptides in cancerous  
 CC tumours and for screening for differentially expressed genes using  
 CC microarray technology. The present sequence represents a human PRO  
 CC protein of the invention.

XX Sequence 274 AA:

Query Match 73.6%; Score 1381; DB 23; Length 274;  
 Best Local Similarity 99.3%; Pred. No. 4,4e-112;  
 Matches 272; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MASMAVILTMALALISAFSATOARKGFMDYFSQTSQDKGRVQIHQOKMAREPATLKDSL 60  
 DB 1 MASMAVILTMALALISAFSATOARKGFMDYFSQTSQDKGRVQIHQOKMAREPATLKDSL 60  
 QY 61 EDDLNNMNFLEKRLPLSGSEAPRLPDQPVGMKRLQDELEEVKARLQPYMAEAHELVCVM 120  
 DB 61 EDDLNNMNFLEKRLPLSGSEAPRLPDQPVGMKRLQDELEEVKARLQPYMAEAHELVCVM 120  
 QY 121 NLEGLRQOLKPYTMDIMEQVALRVQELQVQLRVYVGDHTKAOLIGCVDFWALLQGLQSRV 180  
 DB 121 NLEGLRQOLKPYTMDIMEQVALRVQELQVQLRVYVGDHTKAOLIGCVDFWALLQGLQSRV 180  
 QY 181 VHTGFRFKELFHPYAKSLVSGIGRHVQELHRSVAHPAPASPARLSHCVOVLSNKLITLAKK 240  
 DB 181 VHTGFRFKELFHPYAKSLVSGIGRHVQELHRSVAHPAPASPARLSHCVOVLSNKLITLAKK 240  
 QY 241 ALHARTIQNLQDLRELSRAFAFGTGEAGPDP 274  
 DB 241 ALHARTIQNLQDLRELSRAFAFGTGEAGPDP 274

RESULT 14  
 AAY54561  
 ID AAY54561 standard; Protein; 367 AA.

XX AAY54561;  
 AC  
 XX  
 DT 25-APR-2000 (first entry)  
 XX  
 DE Amino acid sequence of the rat RAP3 protein.  
 XX  
 KW RAP3; regeneration association protein 3; liver regeneration;  
 XX liver proliferation.  
 OS Rattus sp.

XX  
 PN WO200003013-A2.  
 XX  
 PD 20-JAN-2000.  
 XX  
 XX 12-JUL-1999; 99WO-EP04938.  
 PF  
 XX 10-JUL-1998; 98EP-0202336.  
 PR  
 XX (AMST-) AMSTERDAM MOLECULAR THEKAPEUTICS NV.  
 PA  
 XX  
 PI Chamulleau RAFM, Groenink M, Van Der Vliet HN, Leeuwater ACJ;  
 XX  
 DR WPI: 2000-147615/13.  
 XX N-PSDB; AA45872, AA45873.  
 XX  
 PT Isolated RAP3 gene, protein and antibody useful for diagnosing liver  
 PT regeneration and/or cell proliferation -  
 PS Claim 6; Fig 2; 42pp: English.

XX The present sequence represents a rat RAP3 (regeneration association  
 CC protein 3). The RAP3 cDNA sequence was isolated from a rat liver cDNA  
 CC library. The RAP3 gene is involved in regeneration processes of the  
 CC liver. The RAP3 gene was found to be upregulated 6 hours after partial  
 CC hepatectomy, after which it was downregulated. The RAP3 cDNA sequence  
 CC is useful as a source for PCR primers and probes, which are useful  
 CC for detecting nucleotide sequences in a source material. The RAP3  
 CC sequence is useful as a marker of liver proliferation. The RAP3  
 CC protein is useful for the diagnosis of liver regeneration and liver  
 CC cell proliferation. RAP3 antibodies, PCR primers and probes are useful  
 CC for detecting the occurrence of liver cell proliferation in a patient.  
 CC The RAP3 protein is also useful for enhancing the growth of  
 CC regeneration of liver tissue comprising treating the liver tissue  
 CC such as extracorporeal or intracorporeal.

XX Sequence 367 AA:

Query Match 68.9%; Score 1293; DB 21; Length 367;  
 Best Local Similarity 71.8%; Pred. No. 3e-104;  
 Matches 260; Conservative 32; Mismatches 62; Indels 8; Gaps 4;

QY 4 MAVITMALALISAFSATOARKGFMDYFSQTSQDKGRVQIHQOKMAREPATLKDSL 63  
 DB 1 MAVITMALALISAFSATOARKGFMDYFSQTSQDKGRVQIHQOKMAREPATLKDSL 63  
 QY 64 LNNMNFLEKRLPLSGSEAPRLPDQPVGMKRLQDELEEVKARLQPYMAEAHELVCVM 121  
 DB 64 LNNMNFLEKRLPLSGSEAPRLPDQPVGMKRLQDELEEVKARLQPYMAEAHELVCVM 121  
 QY 122 LNEGLRQOLKPYTMDIMEQVALRVQELQVQLRVYVGDHTKAOLIGCVDFWALLQGLQSRV 181  
 DB 122 LNEGLRQOLKPYTMDIMEQVALRVQELQVQLRVYVGDHTKAOLIGCVDFWALLQGLQSRV 181  
 QY 181 VHTGFRFKELFHPYAKSLVSGIGRHVQELHRSVAHPAPASPARLSHCVOVLSNKLITLAKK 241  
 DB 181 VHTGFRFKELFHPYAKSLVSGIGRHVQELHRSVAHPAPASPARLSHCVOVLSNKLITLAKK 241  
 QY 242 LHAARTIQNLQDLRELSRAFAFGTGEAGPDP 274  
 DB 242 LHAARTIQNLQDLRELSRAFAFGTGEAGPDP 274  
 QY 300 TRALIDETEEVQOOLAPPPGSAFAPEFQRTYSGVLSKLAARDLDMEDITHSLHDCG 359  
 DB 300 TRALIDETEEVQOOLAPPPGSAFAPEFQRTYSGVLSKLAARDLDMEDITHSLHDCG 359  
 QY 360 HS 361  
 DB 357 HS 358

RESULT 15  
 AAY67399





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: January 23, 2003, 17:42:47 ; Search time 22 Seconds  
(without alignments)  
489,490 Million cell updates/sec

Title: US-09-842-364-3

Perfect score: 1877  
Sequence: 1 MASMAVLTWALALALSAFSA.....LMEDITHSLHDGSHLGD 366

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1877	100.0	366	US-09-750-580-3	Sequence 3, Appli
2	200	10.7	267	US-07-959-946-3	Sequence 3, Appli
3	200	10.7	267	US-08-333-577-3	Sequence 3, Appli
4	200	10.7	267	US-08-952-796-2	Sequence 2, Appli
5	200	10.7	267	PCT-US92-08634-3	Sequence 3, Appli
6	183.5	9.8	264	US-08-448-606-6	Sequence 6, Appli
7	178	9.5	200	US-08-952-796-15	Sequence 15, Appli
8	162	8.6	317	US-08-949-155-6	Sequence 6, Appli
9	162	8.6	317	US-09-819-964-6	Sequence 6, Appli
10	149	7.9	317	US-07-709-949-2	Sequence 2, Appli
11	128.5	6.8	2101	US-08-456-380-4	Sequence 4, Appli
12	128.5	6.8	2101	US-08-470-950-4	Sequence 4, Appli
13	128.5	6.8	2101	US-08-467-781-4	Sequence 4, Appli
14	128.5	6.8	2101	US-08-195-487-4	Sequence 4, Appli
15	128.5	6.8	2101	US-08-483-924-4	Sequence 4, Appli
16	128.5	6.8	2101	US-09-452-294-1	Sequence 1, Appli
17	128.5	6.8	2101	PCT-US93-06160-4	Sequence 4, Appli
18	123	6.6	576	US-08-533-306A-2	Sequence 2, Appli
19	123	6.6	576	US-08-742-923A-2	Sequence 2, Appli
20	123	6.6	816	US-08-533-306A-6	Sequence 6, Appli
21	123	6.6	816	US-08-742-923A-6	Sequence 6, Appli
22	123	6.6	885	US-08-533-306A-4	Sequence 4, Appli
23	123	6.6	885	US-08-742-923A-4	Sequence 4, Appli
24	121.5	6.5	803	US-09-154-750A-85	Sequence 85, Appli
25	120	6.4	1618	US-07-853-913-4	Sequence 4, Appli
26	116	6.2	1388	US-09-572-191-2	Sequence 2, Appli
27	116	6.2	1388	US-09-723-262-2	Sequence 2, Appli

28	116	6.2	1388	4	US-09-723-219-2	Sequence 2, Appli
29	115	6.1	1090	4	US-09-085-1998-5	Sequence 5, Appli
30	115	6.1	1939	4	US-09-310-187A-1	Sequence 1, Appli
31	115	6.1	2293	4	US-09-368-590-2	Sequence 2, Appli
32	114.5	6.1	1886	4	US-08-938-105-3	Sequence 3, Appli
33	114	6.1	220	2	US-08-726-306A-29	Sequence 29, Appli
34	114	6.1	914	4	US-09-085-1998-4	Sequence 4, Appli
35	113	6.0	566	2	US-08-533-669A-8	Sequence 8, Appli
36	113	6.0	566	2	US-08-511-872-2	Sequence 2, Appli
37	113	6.0	566	4	US-09-183-861-8	Sequence 8, Appli
38	113	6.0	566	4	US-09-022-765-8	Sequence 8, Appli
39	113	6.0	2482	1	US-08-328-254-6	Sequence 6, Appli
40	112.5	6.0	1068	4	US-09-085-1998-11	Sequence 11, Appli
41	112	6.0	3248	5	US-08-353-700-1	Sequence 1, Appli
42	112	6.0	3248	5	PCT-US95-16216-1	Sequence 1, Appli
43	110	5.9	64	2	US-08-282-870-1	Sequence 1, Appli
44	107.5	5.7	331	4	US-08-695-458-2	Sequence 2, Appli
45	107.5	5.7	599	3	US-08-556-419-22	Sequence 22, Appli

## ALIGNMENTS

```

RESULT 1
US-09-750-580-3
Sequence 3, Application US/09750580
Patent No. 6455280
GENERAL INFORMATION:
APPLICANT: Yen, Frances
APPLICANT: Denison, Blake
APPLICANT: Bour, Barbara
APPLICANT: Bihain, Bernard
APPLICANT: Dumas Milne Edwards, Jean-Baptiste
APPLICANT: Duclert, Aymeric
APPLICANT: Bougueleret, Lydie
APPLICANT: Ebbels-Reed, Dana
APPLICANT: Saller-Cid, Luisa
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING NEOPLASTIC CELL GROWTH
FILE REFERENCE: 89, US2, CIP
CURRENT APPLICATION NUMBER: US/09/750, 580
CURRENT FILING DATE: 2000-12-28
PRIOR APPLICATION NUMBER: US 09/599, 362
PRIOR FILING DATE: 2000-06-21
PRIOR APPLICATION NUMBER: PCT/IB00/0101
PRIOR FILING DATE: 2000-06-21
PRIOR APPLICATION NUMBER: PCT/IB99/02058
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: US 49/469/099
PRIOR FILING DATE: 1999-12-21
PRIOR APPLICATION NUMBER: US 60/113, 686
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: US 60/141, 032
PRIOR FILING DATE: 1999-06-25
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Patent.pm
SEQ ID NO 3
LENGTH: 366
TYPE: PRT
ORGANISM: Homo sapiens
US-09-750-580-3
Query Match 100.0%; Score 1877; DB 4; Length 366;
Best Local Similarity 100.0%; Pred. No. 4.4e-161;
Matches 366; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MASMAVLTWALALALSAFSAATQARKGFWDYFSQTSQDGRVEQIHOQKMAREPTLDSL 60
DB 1 MASMAVLTWALALALSAFSAATQARKGFWDYFSQTSQDGRVEQIHOQKMAREPTLDSL 60
QY 61 EODLNNMKKFLKRLPLSGSEAPRLPDPVGMRRLOLELEEVYAKRQPYMAEHELVGW 120
DB 61 EODLNNMKKFLKRLPLSGSEAPRLPDPVGMRRLOLELEEVYAKRQPYMAEHELVGW 120

```

```

QY 121 NLEGLROQLKPYTMDLMEQVALRVQELQELRVGVEDTKAOLLGVDAMALLQGLQSRV 180
    |||||
DB 121 NLEGLROQLKPYTMDLMEQVALRVQELQELRVGVEDTKAOLLGVDAMALLQGLQSRV 180
QY 181 VNHGRKRELPHPAESLYSGICRHVOELHRSVAPHAPASPARLSRCVOVLSRKLTLLAK 240
    |||||
DB 181 VNHGRKRELPHPAESLYSGICRHVOELHRSVAPHAPASPARLSRCVOVLSRKLTLLAK 240
QY 241 ALAARIDONLDQRELSRAFGCTGTEGAGPDPOMLSEEVORLQAFRODTYQIAAFT 300
    |||||
DB 241 ALAARIDONLDQRELSRAFGCTGTEGAGPDPOMLSEEVORLQAFRODTYQIAAFT 300
QY 301 RAIDQETEEVOQLAPPPGHSAPAFEFQOTDSGKVLKLOARLDLDMEDITHSLHDOGH 360
    |||||
DB 301 RAIDQETEEVOQLAPPPGHSAPAFEFQOTDSGKVLKLOARLDLDMEDITHSLHDOGH 360
QY 361 SHLQDP 366
    |||||
DB 361 SHLQDP 366

```

```

RESULT 2
US-07-959-946-3
: Sequence 3, Application US/07959946
: Patent No. 5408038
: GENERAL INFORMATION:
: APPLICANT: Smith, Richard K.
: APPLICANT: Koduri, Raju
: APPLICANT: Young, Stephen G.
: APPLICANT: Witzlum, Joseph L.
: APPLICANT: Curtiss, Linda K.
: TITLE OF INVENTION: Lipoprotein Assays Using Antibodies to a
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Dressler, Goldsmith, Shore, Sutker &
: ADDRESSEE: Milnamow, Ltd.
: STREET: 180 No. 5408038th Stetson, Suite 4700
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60601
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/959,946
: FILING DATE: 19921008
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/901,706
: FILING DATE: 18-JUN-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Gamson, Edward P.
: REGISTRATION NUMBER: 29,381
: REFERENCE/DOCKET NUMBER:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (312)616-5400
: TELEFAX: (312)616-5460
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 267 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-07-959-946-3

```

Query Match 10.7%; Score 200; DB 1; Length 267;  
 Best Local Similarity 22.2%; Pred. No. 2.6e-10;  
 Matches 74; Conservative 61; Mismatches 106; Indels 92; Gaps 10;

```

QY 5 AAVLTWALLALSAFSAATQARKGFMD-----YFSQTSQDKGR--VEQI 45
    |||||
DB 3 AAVLTWALLALSAFSAATQARKGFMD-----YFSQTSQDKGR--VEQI 45
QY 46 OOKMARE-PATLKSDLEODLNMMKFLKRLPLSSSEAPRLQDQVYGNARQLOEELIEYK 104
    : : :
DB 59 GSALQKQNLTKLLDMWDVSTFESKLRQDLGFTVDFEFDNLKEKTEGRLQENSKLIEYK 118
QY 105 ARLQPYMAEHLVGMNLYGRLQOLKPYTMDLMEQVALRVQELQELRVGVEDTKAOLLG 164
    |||||
DB 119 AKVOPLRDDPQKQDEBEMELVROKVEPLRAELQECAROKLHELQKLSPLGEMR----- 173
QY 165 GVDAMALLQGLQSRVNHHTGRFKELPHPAESLYSGICRHVOELHRSVAPHAPASPARL 224
    |||||
DB 174 --DRARA-----HYDALRTHLAAPYSDFLRQL 198
QY 225 SNCVOVLSKRLTLKAKALHARIDONLDQRELSRAFGCTGTEGAGPDPOMLSEEVOR 284
    : : :
DB 199 AARLEALKENGARLAELVHAKATEHLSTLERAKPAL-----EDLRQG 241
QY 285 ----LQAFRODTYQIAAFTRAIDQETEEVOQ 313
    |||||
DB 242 LLPVLESFR-----VSFLSALBEVTKKLTQ 267

```

```

RESULT 3
US-08-333-577-3
: Sequence 3, Application US/08333577
: Patent No. 5786206
: GENERAL INFORMATION:
: APPLICANT: Smith, Richard K.
: APPLICANT: Koduri, Raju
: APPLICANT: Young, Stephen G.
: APPLICANT: Witzlum, Joseph L.
: APPLICANT: Curtiss, Linda K.
: TITLE OF INVENTION: Lipoprotein Assays Using Antibodies to a
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Dressler, Goldsmith, Shore, Sutker &
: ADDRESSEE: Milnamow, Ltd.
: STREET: 180 No. 5786206th Stetson, Suite 4700
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60601
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/333,577
: FILING DATE:
: CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
: NAME: Gamson, Edward P.
: REGISTRATION NUMBER: 29,381
: REFERENCE/DOCKET NUMBER: SCRF 234.0
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (312)616-5400
: TELEFAX: (312)616-5460
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 267 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-333-577-3

```

Query Match 10.7%; Score 200; DB 1; Length 267;  
 Best Local Similarity 22.2%; Pred. No. 2.6e-10;  
 Matches 74; Conservative 61; Mismatches 106; Indels 92; Gaps 10;

```

OY 5 AAVLTWALALSAFSAATQARKGFWD-----YFSQTSQDGKGR--VEQIH 45
    ||||| : : : ||| ||
DB 3 AAVLTAVLFL---TGSQARH-FWQODEPQSPMDRVKDLATVYVDVAKDSGRDVSQFE 58
OY 46 OOKMARE-PATLKDSLEODLNMMKFLKRLPLSGSEAPRLPDPVGNRRLOLELEEVK 104
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 59 GSALCKQNLNKLDMWDVSTFSFKRLRQGLGVTQGFMDNLEKEETGEGRQENSKDLEEVK 118
OY 105 ARIQYMAVAHELVGMNLEGLRQOLKPYTMDIMEQVALRVQELQOLRVGCHDTKAQILLG 164
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 119 AKVQPYLDDFOKKMOEEMELTKQKVEPLRAELQEGAROKLHLEQKLSPLGEMR----- 173
OY 165 GVDEAMALLQGLQSRVHVHTGFKELFHPYAESLVSGIGRHVOELHRSVAPHPASPAPRL 224
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 174 --DRARA-----HYDALRTHILAPYSDELRL 198
OY 225 SRCVOVLSRKLTAKAKALHARIQNLIDQLRELSAFACTGTEGAGDPQMLSEVBQR 284
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 199 AARLEALKENGARLAEXHAKATEHLSTLSEKAKPAL-----EDLRQG 241
OY 285 ----LQAFKQDTYDQIAAFTRAIDQETEEVQOQ 313
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 242 LLPVLESFK-----VSFLSALLEYTKKLNTQ 267

```

## RESULT 4

```

US-08-952-796-2
: Sequence 2, Application US/08952796
: Patent No. 6258596

```

## GENERAL INFORMATION:

```

: APPLICANT: BENOTTI, Patrick
: APPLICANT: BRUCKERT, Eric
: APPLICANT: DENEFLE, Patrice
: APPLICANT: DUBERGER, Nicolas
: APPLICANT: FRUCHART, Jean-Charles
: APPLICANT: LUC, Gerald
: APPLICANT: TURPIN, Gerard
: APPLICANT: ASSMANN, Gerd
: TITLE OF INVENTION: NOVEL VARIANTS OF APOLIPOPROTEIN A-1
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Rhone-Poulenc Korer Inc.
: STREET: 500 Arcola Road, Mailstop 3C43
: CITY: Collegeville
: STATE: PA
: COUNTRY: USA

```

```

: ZIP: 19426
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC Compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/952.796
: FILING DATE:
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 95/06061
: FILING DATE: 22-MAY-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO FR96/00747
: FILING DATE: 20-MAY-1996

```

## ATTORNEY/AGENT INFORMATION:

```

: NAME: Fehner Esq., Paul F.
: REGISTRATION NUMBER: 35,135
: REFERENCE/DOCKET NUMBER: ST95031-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (610) 454-3839
: TELEFAX: (610) 454-3808
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:

```

```

: LENGTH: 267 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-952-796-2

```

```

Query Match 10.7%, Score 200, DB 4; length 267;
Best Local Similarity 22.2%, pred. No. 2,6e-10;
Matches 74; Conservative 61; Mismatches 106; Indels 92; Gaps 10;

```

```

OY 5 AAVLTWALALSAFSAATQARKGFWD-----YFSQTSQDGKGR--VEQIH 45
    ||||| : : : ||| ||
DB 3 AAVLTAVLFL---TGSQARH-FWQODEPQSPMDRVKDLATVYVDVAKDSGRDVSQFE 58
OY 46 OOKMARE-PATLKDSLEODLNMMKFLKRLPLSGSEAPRLPDPVGNRRLOLELEEVK 104
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 59 GSALCKQNLNKLDMWDVSTFSFKRLRQGLGVTQGFMDNLEKEETGEGRQENSKDLEEVK 118
OY 105 ARIQYMAVAHELVGMNLEGLRQOLKPYTMDIMEQVALRVQELQOLRVGCHDTKAQILLG 164
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 119 AKVQPYLDDFOKKMOEEMELTKQKVEPLRAELQEGAROKLHLEQKLSPLGEMR----- 173
OY 165 GVDEAMALLQGLQSRVHVHTGFKELFHPYAESLVSGIGRHVOELHRSVAPHPASPAPRL 224
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 174 --DRARA-----HYDALRTHILAPYSDELRL 198
OY 225 SRCVOVLSRKLTAKAKALHARIQNLIDQLRELSAFACTGTEGAGDPQMLSEVBQR 284
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 199 AARLEALKENGARLAEXHAKATEHLSTLSEKAKPAL-----EDLRQG 241
OY 285 ----LQAFKQDTYDQIAAFTRAIDQETEEVQOQ 313
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 242 LLPVLESFK-----VSFLSALLEYTKKLNTQ 267

```

## RESULT 5

```

PCR-US92-08634-3
: Sequence 3, Application PC/TUS9208634

```

## GENERAL INFORMATION:

```

: APPLICANT: Smith, Richard K.
: APPLICANT: Koduri, Raju
: APPLICANT: Young, Stephen G.
: APPLICANT: Witzum, Joseph L.
: APPLICANT: Curtis, Linda K.
: TITLE OF INVENTION: Lipoprotein Assays Using Antibodies to a
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dressler, Goldsmith, Shore, Sulker &
: STREET: 180 North Sletson, Suite 4700
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA

```

```

: ZIP: 60601
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC Compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US92/08634
: FILING DATE: 19921009
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/901,706
: FILING DATE: 18-JUN-1992

```

## ATTORNEY/AGENT INFORMATION:

```

: NAME: Gamson, Edward P.
: REGISTRATION NUMBER: 29,381
: REFERENCE/DOCKET NUMBER:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (312)616-5400

```

TELEFAX: (312)616-5460  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 267 amino acids  
 TYPE: AMINO ACID  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 PCT-US92-08634-3

Query Match 10.7%; Score 200; DB 5; length 267;  
 Best local Similarity 22.2%; Pred. No. 2.6e-10;  
 Matches 74; Conservative 61; Mismatches 106; Indels 92; Gaps 10;

5 AAVLTMAALLAFSAFATQARKGFWD-----YFSQTSQDKGR--VEQIIT 45  
 3 AAVLTIAVIFL---TGSQAKH-FWQODDEPQSPMDVKVATLVYVYVKDSCGDVYSQF 58  
 46 GQKMAE-PATIKDSLEODLNMMKFEKLRPLSGSEARLPDPDVGRRLOLEEVK 104  
 59 GVALGKQLMLKLDNMDSVTSTFSKREQLGPVQEFMDNLEKEGRLQEMSKDLEEVK 118  
 105 ARLQPMMAHFLVGMNLEGLKQOLKPYTMDLMEQVALVQELQRLVYVGEYTKAQLIG 164  
 119 AKVQPLDGFQKQWQEMELVYKVEPLRAELQEGAROKLHLEQKLSPLGEMR----- 173  
 165 GVDEAMALLQGLQSRVYHHTGFKELFHPYAESLYSGIRHVOELHRSVAPHPAPAPARL 224  
 174 ---DRAA-----HVALRTHLAPYSDELRLQRL 198  
 225 SRGVOLSKRLKAKALARIQONIDQLRELSRAFACTGTEEGAGPPQMLSEEVRR 284  
 199 ARLLELKGARLALEYHAKATEHLSTLSEKAKPAL-----EDLRQG 241  
 285 ----LQAFRODTYLOIAAFRAIDQETEVQOQ 313  
 242 LRPVLESFR-----VSFLSALLETYTKRLNTO 267

RESULT 6  
 US-08-448-606-6  
 Sequence 6, Application US/08448606  
 Patent No. 5721114

GENERAL INFORMATION:  
 APPLICANT: Adrahams n, Lars  
 APPLICANT: Holmgren, Erik  
 APPLICANT: Kalden n, Christina  
 APPLICANT: Lake, Mats  
 APPLICANT: Mikaelsson, SA  
 APPLICANT: Seijltz, Torsten  
 TITLE OF INVENTION: Expression System For Producing  
 TITLE OF SEQUENCES: Apolipoprotein AI-M  
 NUMBER OF SEQUENCES: 7  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Pollock, Vande Sande & Priddy  
 STREET: 1990 M Street, N.W., Suite 800  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: U.S.  
 ZIP: 20036  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/448,606  
 FILING DATE: 25-AUG-1995  
 CLASSIFICATION: 436  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/SE93/01061  
 FILING DATE: 09-DEC-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: SE 9203753-0

FILING DATE: 11-DEC-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Americk, Burton A.  
 REGISTRATION NUMBER: 24,852  
 REFERENCE/DOCKET NUMBER: 0151/00121  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202)331-7111  
 TELEFAX: (202) 293-6229  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 264 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-448-606-6

Query Match 9.8%; Score 183.5; DB 1; length 264;  
 Best local Similarity 22.8%; Pred. No. 7.7e-09;  
 Matches 62; Conservative 55; Mismatches 122; Indels 33; Gaps 6;

1 MASMAVLTMAAL-LLSATSATQARKGFWD-----YFSQTSQDKGR--VEQIHOOKMA 50  
 1 MKRYAIAIAVALAGFATVNADEPQSPMDRVKDIATVYVDVVKDSCRDVYSQFEGSALG 60  
 51 RE-PATIKDSLEODLNMMKFEKLRPLSGSEARLPDPDVGRRLOLEEVKARLP 109  
 61 KQNLKLDNMDSVTSTFSKREQLGPVQEFMDNLEKEGRLQEMSKDLEEVKAVQ 120  
 121 YLDDFQKQWQEMELVYKVEPLRAELQEGAROKLHLEQKLSPLGEMR-----DRA 173  
 110 YMAEHLVGMNLEGLRQOLKPYTMDLMEQVALVQELQRLVYVGEYTKAQLIGVDRA 169  
 170 WALQGLQSRVYHHTGFKELFHPYAESLYSGIRHVOELHRSVAPHPAPAPARL 229  
 174 RAHVADLRTHLAPYSDELQCLAKRLKLENGCARIAFYHAKATEHLSTL----- 223  
 230 VLSRKLTKAKALARIQONIDQLRELSRAFA 261  
 224 ----LSEKAKPALRDLRLQCLLPVLESFVSWF 250

RESULT 7  
 US-08-952-796-15  
 Sequence 15, Application US/08952796  
 Patent No. 6258596

GENERAL INFORMATION:  
 APPLICANT: Benoit, Patrick  
 APPLICANT: BROCKERT, Eric  
 APPLICANT: DENEFLE, Patrice  
 APPLICANT: DIBERGER, Nicolas  
 APPLICANT: FRUCHART, Jean-Charles  
 APPLICANT: IUC, Gerard  
 APPLICANT: TURPIN, Gerard  
 APPLICANT: ASSMANN, Gerd  
 APPLICANT: FUNKE, Harald  
 TITLE OF INVENTION: NOVEL VARIANTS OF APOLIPOPROTEIN A-I  
 NUMBER OF SEQUENCES: 19  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Rhone-Poulenc Rorer Inc.  
 STREET: 500 Arcola Road, Mailstop 3C43  
 CITY: Collegeville  
 STATE: PA  
 COUNTRY: USA  
 ZIP: 19426  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/952,796  
 FILING DATE:  
 CLASSIFICATION: 424



FILING DATE: 28-Mar-2001  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/949,155  
 FILING DATE: <Unknown>  
 APPLICATION NUMBER: US 60/046,094  
 FILING DATE: 09-MAY-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Hilder, David M.  
 REGISTRATION NUMBER: 41,071  
 REFERENCE/DOCKET NUMBER: TAMK:177  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (512) 418-3000  
 TELEFAX: (713) 789-2679  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 317 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
 US-09-819-964-6

Query Match 8.6%; Score 162; DB 4; Length 317;  
 Best Local Similarity 20.8%; Pred. No. 8.5e-07;  
 Matches 63; Conservative 66; Mismatches 112; Indels 62; Gaps 9;

22 QARKGFMDYTSQSGDKGRV-EOIHQKMAREPATLKDSLEODLNKFKLEKLR---P 76  
 45 QALGRMDYLRWQSLSDQVEELSTKYQE---LTELIEESKKEKAVREBELQDLP 101  
 77 LSSSEARLPDPDVGMRKLOELEEYKARLOPYMAEHLVGMNLEGLRQOQKPYTMDL 136  
 102 VTETQARLSKELOAQARAGADEMDYRNLVYRSEVHNMLOQTEELSRSLASHRLK 161  
 137 MPQVALRVOELQOLRVYGEDTKAQLLQGVDEAMALLOQSRVHHTGFKELFHPYAE 196  
 162 RKKLRTELQOKRLAVY---QAGLRKGRSVSLREKLPVEO-GRLR-----AA 210  
 197 SLVSGIGRHOELHRSVAPAPASPARLSRCVOYLSRKLTLKAKALHARIQONLDQREE 256  
 211 TLSTRAGQPLRER-----AAWGCOKLRCGRLEMGSRTRDLDEMREQ 252  
 257 LSAFAGTGEBCAGPDPOMLSEEVORLOAFQODPYLQIAATFRAI---DOETEEVQO 312  
 253 L-----EVRKVEEGSQSLRLQAEHFALLKGFEPLEVEDIR 291  
 313 QLA 315  
 292 QWA 294

RESULT 10  
 US-07-709-949-2  
 Sequence 2, Application US/07709949  
 Patent No. 5472858  
 GENERAL INFORMATION:  
 APPLICANT: Attie, Alan D.  
 APPLICANT: Gretch, Daniel G.  
 APPLICANT: Sturley, Stephen L.  
 APPLICANT: Beckage, Nancy E.  
 TITLE OF INVENTION: Production of Recombinant Proteins in  
 TITLE OF INVENTION: Insect Larvae  
 NUMBER OF SEQUENCES: 2  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Quarles & Brady  
 STREET: P.O. Box 2113  
 CITY: Madison  
 STATE: WI  
 COUNTRY: USA  
 ZIP: 53701-2113  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07709,949  
 FILING DATE: 19910604  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Seay, Nicholas J.  
 REGISTRATION NUMBER: 27,386  
 REFERENCE/DOCKET NUMBER: 9629691801  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (608) 251-5000  
 TELEFAX: (608) 251-9166  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 317 amino acids  
 TYPE: AMINO ACID  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-07-709-949-2

Query Match 7.9%; Score 149; DB 1; Length 317;  
 Best Local Similarity 22.0%; Pred. No. 1.3e-05;  
 Matches 69; Conservative 59; Mismatches 117; Indels 68; Gaps 10;

10 MALALLSASFATQARKGFMDY--FSQTSQDKGRVEOIHQKMAREPATLKDSLEODLNK 67  
 44 WEAL-----GRFMDYLRWQSLSDQVEELSSQVQOELRALDETMKELKAY 92  
 68 NKFL-EKLRPLSSSEARLPDPDVGMRKLOELEEYKARLOPYMAEHLVGMNLEGLR 126  
 93 KSELEOLTPEAEETRLARLSKELOTAQARAGADEMDYCGRLVYRGVQAMLGOSTEELR 152  
 127 QOLKPYTMDMEQVALRVOELQOLRVYGEDTKAQLLQGVDEAMALLOQSRVHHTG 186  
 153 VLSASHRLKRLKRLRDPDQOKRLAVY---QAGAREGAE-----KGL-SAIREKLP 201  
 187 FKELFHPYAESLVSGIGRHOELHRSVAPAPASPARLSRCVOYLSRKLTLKAKALHARI 246  
 202 LVEGGRVATVGSLSAQPLQER-----AQAMGERLARMBEGSRT 243  
 247 QONLDQREELSRFACTGTEBCAGPDPOMLSEEVORLOAFQODPYLQIAATFRAI--- 303  
 244 RDRDEVEKEQVA-----EVRKLEEQAOQIRLQAEAFQARLKSW 282  
 304 -DOETEEVQOOLA 315  
 283 FHPYLVHDMQKQWA 295

RESULT 11  
 US-08-466-390-4  
 Sequence 4, Application US/08466390  
 Patent No. 5686562  
 GENERAL INFORMATION:  
 APPLICANT: TOUKATLY, GARY  
 APPLICANT: LIDGARD, GRAHAM P.  
 TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
 TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
 NUMBER OF SEQUENCES: 6  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
 STREET: 125 HIGH STREET  
 CITY: BOSTON  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02110  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,390
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-466-390-4

```

```

Query Match 6.8%; Score 128.5; DB 1; Length 2101;
Best local Similarity 22.0%; Pred. No. 0.013;
Matches 81; Conservative 57; Mismatches 131; Indels 99; Gaps 15;

QY 37 DKGRVEQIHQOKMAREPATLKDSLQDLNNKFLKRLPLSGSEAPRL----- 85
DB 453 ERGHFEERKQOL-----SLITDLOSSISNLSQAKKELEQASQAHGARTNOVASLTSEL 507
QY 86 -----PODPVGMROLOLELEVKARLOPYMAEAHVLGCMNLGCLROOLKPYTMD 135
DB 508 TTLNATIQOQDDELGLKQAKERQOLAOTLQOQEQAS-----QGLRHQVEQLSSS 559
QY 136 L--MEQVALRYVOLE-----QLRVGEDTAAOLGVDAMALIGIOS-----R 179
DB 560 LKQEQOLKEVAEQEATRODHAQOLATAEERESLRLER--DAALKQLEALEKEKAAKLE 618
QY 180 VVHTGRKELFHPYAESLVSGIGHVOELHRSVAAPHASPAPLSRCV----- 228
DB 619 ILQOQLOVANEARDASQTSVTOQREKALSRKY-----EELQACVETAROBQHEAQ 670
QY 229 -QVLSRLTLKAKALHA-----RIOQNI.DQLRELSRAFACTGTEGACPDPMUSEEVRO 283
DB 671 AQVALELQLSQOKALEKERVAKQEKDQLOALKESELKVTGKS-----LEEKR 724
QY 284 RLQAF--RODTYLOIAAFTRAIQDETEVEQOQLAPPPGHSAPAFPEQOOTSQVLSKIQ 341
DB 725 AADALEEQORCISELKATRSJL-----VEQ-----HKRRKELFEERAGR--KGLE 768
QY 342 ARLDLME 349
DB 769 ARLLQGE 776

```

```

RESULT 12
US-08-470-950-4
Sequence 4, Application US/08470950
Patent No. 5698439
GENERAL INFORMATION:
APPLICANT: TONKATLY, GARY
APPLICANT: LIDGARD, GRAHAM P
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
NUMBER OF SEQUENCES: 6
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: 125 HIGH STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,950
FILING DATE: 06-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-470-950-4

```

```

Query Match 6.8%; Score 128.5; DB 1; Length 2101;
Best local Similarity 22.0%; Pred. No. 0.013;
Matches 81; Conservative 57; Mismatches 131; Indels 99; Gaps 15;

QY 37 DKGRVEQIHQOKMAREPATLKDSLQDLNNKFLKRLPLSGSEAPRL----- 85
DB 453 ERGHFEERKQOL-----SLITDLOSSISNLSQAKKELEQASQAHGARTNOVASLTSEL 507
QY 86 -----PODPVGMROLOLELEVKARLOPYMAEAHVLGCMNLGCLROOLKPYTMD 135
DB 508 TTLNATIQOQDDELGLKQAKERQOLAOTLQOQEQAS-----QGLRHQVEQLSSS 559
QY 136 L--MEQVALRYVOLE-----QLRVGEDTAAOLGVDAMALIGIOS-----R 179
DB 560 LKQEQOLKEVAEQEATRODHAQOLATAEERESLRLER--DAALKQLEALEKEKAAKLE 618
QY 180 VVHTGRKELFHPYAESLVSGIGHVOELHRSVAAPHASPAPLSRCV----- 228
DB 619 ILQOQLOVANEARDASQTSVTOQREKALSRKY-----EELQACVETAROBQHEAQ 670
QY 229 -QVLSRLTLKAKALHA-----RIOQNI.DQLRELSRAFACTGTEGACPDPMUSEEVRO 283
DB 671 AQVALELQLSQOKALEKERVAKQEKDQLOALKESELKVTGKS-----LEEKR 724
QY 284 RLQAF--RODTYLOIAAFTRAIQDETEVEQOQLAPPPGHSAPAFPEQOOTSQVLSKIQ 341
DB 725 AADALEEQORCISELKATRSJL-----VEQ-----HKRRKELFEERAGR--KGLE 768
QY 342 ARLDLME 349
DB 769 ARLLQGE 776

```

```

RESULT 13
US-08-467-781-4
Sequence 4, Application US/08467781
Patent No. 5780596
GENERAL INFORMATION:
APPLICANT: TONKATLY, GARY
APPLICANT: LIDGARD, GRAHAM P
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
NUMBER OF SEQUENCES: 6
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: 125 HIGH STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```



```

1 OPERATING SYSTEM: PC-DOS/MS-DOS
2 SOFTWARE: Patent In Release #1.0, Version #1.25
3 CURRENT APPLICATION DATA:
4 APPLICATION NUMBER: US/08/467,781
5 FILING DATE: 06-JUN-1995
6 CLASSIFICATION: 424
7 ATTORNEY/AGENT INFORMATION:
8 NAME: PITCHER ESO, EDMUND R
9 REGISTRATION NUMBER: 27,829
10 REFERENCE/DOCKET NUMBER: MTP-013
11 TELECOMMUNICATION INFORMATION:
12 TELEPHONE: (617) 248-7000
13 TELEFAX: (617) 248-7100
14 INFORMATION FOR SEQ. ID NO.: 4:
15 SEQUENCE CHARACTERISTICS:
16 LENGTH: 2101 amino acids
17 TYPE: amino acid
18 TOPOLOGY: linear
19 MOLECULE TYPE: protein
20 OS-08-467-781-4

```

Query Match	6.8%	Score 128.5;	DB 1;	Length 2101;
Best local Similarity	22.0%	Pred. No. 0.013;		
Matches	81; Conservative	57; Mismatches	131; Indels	99; Gaps 15

[illegible]

```

1      RESULT 14
2      US-08-195-487-4
3      :
4      : Sequence 4, Application US/08195487
5      :
6      : Patent No. 5783403
7      :
8      : GENERAL INFORMATION:
9      :
10     : APPLICANT: TOKUKITLY, GARY
11     :
12     : APPLICANT: LIDGARD, GRAHAM P
13     :
14     : TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
15     :
16     : TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
17     :
18     : NUMBER OF SEQUENCES: 6
19     :
20     : CORRESPONDENCE ADDRESSES:
21     :
22     : ADDRESSEE: TESTA HORWITZ & THIBEAULT
23     :
24     : STREET: 53 STATE STREET
25     :
26     : CITY: BOSTON
27     :
28     : STATE: MA
29     :
30     : COUNTRY: USA
31     :
32     : ZIP: 02109
33     :
34     : COMPUTER READABLE FORM:
35     :
36     : MEDIUM TYPE: Floppy disk

```

```

1  COMPUTER:  IBM PC compatible
2  OPERATING SYSTEM:  PC-DOS/MS-DOS
3  SOFTWARE:  PatentIn Release #1.0, Version #1.25
4  CURRENT APPLICATION DATA:
5  APPLICATION NUMBER:  US/08/195,487
6  FILING DATE:
7  CLASSIFICATION:  435
8  Prior Application DATA:
9  APPLICATION NUMBER:  US/07/901,701
10 FILING DATE:
11 ATTORNEY/AGENT INFORMATION:
12 NAME:  PITCHER ESO, EDMUND R
13 REGISTRATION NUMBER:  27,829
14 REFERENCE/DOCKET NUMBER:  MTP-013
15 TELECOMMUNICATION INFORMATION:
16 TELEPHONE:  617/248-7000
17 TELEFAX:  617/248-7100
18 INFORMATION FOR SEQ ID NO:  4:
19 SEQUENCE CHARACTERISTICS:
20     LENGTH:  2101 amino acids
21     type:  amino acid
22     TOPOLOGY:  linear
23     MOLECULE TYPE:  protein
24     OS-08-195-487-4

```

Query Match	6.88	Score 128.5	DB 1	Length 2101
Best Local Similarity	22.08	Pred. No. 0.013		
Matches 81, Conservative	57	Mismatches 131	Indels 99	Caps 15

```

QY 37 DKGRVHQHQKMA REBATTLEDSLEODLNNKFKLEKRLPSGSEAPRL----- 85
Db 453 ERGHFEERKQOL-----SSLTIDQSSISNLSQAKKELEQASQAHGRKRLTAQVASTLSEL 507
QY 86 -----PODPVGNHRKQLOELEEYKARLOPYMAEAHELVGWNLGRLQOLKPYTMD 135
Db 508 TTNATNIOODQDELGAQKQAKKEQAOALQOTLQOQEQAS-----QGRHQYVEOLSSS 559
QY 136 L--MEQVALRVOQLOE-----QLRVGEDTQAOLLGVDENKALLQGLQS-----R 179
Db 560 LKQKEQOLKEVAEKOENTRODHAQOOLATAAEERFASLRER--DALKOLLEALEKEKAkle 618
QY 180 VVHHITGRKELFHFYAESLSVSGIRGHVOTLHRSVAPHAPASPARLSRCV----- 228
Db 619 ILQOOLQVANAARBSAOTSVTAQAREKAELSRK-----EELQACVETARQEOHQA 670
QY 229 -QVLSRKLTLLKAKALHA---RIOONLDOLREELSRFACTGTEEGAGPDOMLSEEVRO 283
Db 671 AQVAELELOLRESEQOKATERERVAKOERQLOEQOLKESLKYTKGS-----LEEKRK 724
QY 284 RLQAF--RODTYLOIAAFTRALIDETEEVVOQLAPPPGHSATAPPELOUTDSKYSKLO 341
Db 725 AADLALEBQOKCSHKLKAFTRSL-----VBO-----HKRERKKELEERKACH--KGLE 768
QY 342 ARLDLWE 349
Db 769 ARLDQGE 776

```

RESULT 15  
US-08-483-924-4  
Sequence 4, Application US/08483924  
Patent No. 5882876  
GENERAL INFORMATION:  
APPLICANT: TOKRATLY, GARY  
APPLICANT: LIDGARD, GRAHAM P  
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
STREET: 125 HIGH STREET  
CITY: BOSTON  
STATE: MA

COUNTRY: USA  
 ZIP: 02110  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/483,924  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: PITCHER ESO, EDMUND R  
 REGISTRATION NUMBER: 27,829  
 REFERENCE/DOCKET NUMBER: MTP-013  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 248-7000  
 TELEFAX: (617) 248-7100  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2101 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-483-924-4

Query Match 6.8%; Score 128.5; DB 2; Length 2101;  
 Best Local Similarity 22.0%; Pred. No. 0.013;  
 Matches 81; Conservative 57; Mismatches 131; Indels 99; Gaps 15;

OY 37 DKGRVEQIHQOMAREPATLQDSLEQDINNKKFLEKLPUSGSPAPRI----- 85  
 Db 453 ERGHFEERKQDL-----SLITDLOSISINLSQAKEELQASQAGCARLTAAVASLTSEL 507  
 OY 86 -----PODPVGMRRLOEELLEVKKARLOPYMAEAHVELVGNLEGTRQOLKPYTMD 135  
 Db 508 TTLNATIQOQDDELAGLKQAKKQAOQLAQTLQOQEOAS-----OQLRHQVEQI,SSS 559  
 OY 136 L--MEQVALRYQELQ-----QLRVGEDTKAQLLGVDAMALLQGLQS-----R 179  
 Db 560 LKQEQQLKEVAEKEQATRODIAQOLATAEEREASLRER-DALQLQLELEKEKAKLE 618  
 OY 180 VVHHTGRKELFHPYAESIYSGICRHVOELHKSVAHPASPAPRLSRCV----- 228  
 Db 619 ILQOQLOVANEARDSAQSVTQAOEREKAELSRKV-----EELQACVETARQDQHEAQ 670  
 OY 229 -QVLSRKLTLKAKALIA-----RIQONLDOLREELSRFAAGTGEAGFDPOMLSEEVRO 283  
 Db 671 AQVAELFIQLKSEQOKATEKKHVAQEKQDLOKQIQALKKSILKVTGKS-----LFEKKR 724  
 OY 284 RLQAF--KQDTYLOIAFTRAIDQETEEVQOQLAPPFGHSFAPEFVOOTDSGKVLKLO 341  
 Db 725 AADALEEQRCISELKAEIRSL-----VEQ-----HKRERKELEEBRAGR--KGLE 768  
 OY 342 ARLDLME 349  
 Db 769 ARLDLGE 776

Search completed: January 23, 2003, 17:47:23  
 Job time : 25 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 23, 2003, 17:43:52 : Search time 40 seconds

(without alignments)  
184.634 Million cell updates/sec

Title: US-09-842-364-3  
Perfect score: 1877  
Sequence: 1 MASMAVLTWALTLASAFSA.....LMEDITHSLDGHSLGDP 366

Scoring table: HIOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 122226 seqs, 20178551 residues

Total number of hits satisfying chosen parameters: 122226

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
Published Applications AA:  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB pep: \*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB pep: \*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB pep: \*  
4: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep: \*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep: \*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep: \*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB pep: \*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep: \*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB pep: \*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB pep: \*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB pep: \*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB pep: \*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB pep: \*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1877	100.0	366	10	US-09-751-877-3 Sequence 3, Appli
2	1877	100.0	366	10	US-09-835-996A-2 Sequence 2, Appli
3	1877	100.0	400	10	US-09-835-996A-41 Sequence 41, Appli
4	1871	99.7	366	10	US-09-800-729-212 Sequence 212, App
5	1858	99.0	363	10	US-09-800-729-111 Sequence 111, App
6	350	18.6	382	9	US-09-987-107-37 Sequence 37, Appli
7	350	18.6	382	10	US-09-800-729-206 Sequence 206, App
8	333	17.7	429	9	US-09-987-107-34 Sequence 34, Appli
9	328	17.5	396	10	US-09-800-729-207 Sequence 207, App
10	327	17.4	396	9	US-09-987-107-33 Sequence 33, Appli
11	315	16.8	401	9	US-09-987-107-36 Sequence 36, Appli
12	311.5	16.6	391	10	US-09-987-107-36 Sequence 36, Appli
13	298.5	15.9	391	10	US-09-800-729-208 Sequence 38, Appli
14	298.5	15.9	395	9	US-09-987-107-35 Sequence 35, Appli
15	200	10.7	267	10	US-09-803-918A-2 Sequence 2, Appli
16	200	10.7	267	10	US-09-803-918A-2 Sequence 2, Appli
17	193	10.3	267	9	US-09-987-107-16 Sequence 16, Appli
18	189	10.1	258	9	US-09-987-107-4 Sequence 4, Appli
19	189	10.1	273	9	US-09-987-107-50 Sequence 50, Appli

20	188	10.0	301	9	US-09-987-107-3	Sequence 3, Appli
21	188	10.0	301	9	US-09-987-107-5	Sequence 5, Appli
22	188	10.0	316	9	US-09-987-107-48	Sequence 48, Appli
23	188	10.0	316	9	US-09-987-107-54	Sequence 54, Appli
24	186	9.9	304	9	US-09-987-107-6	Sequence 6, Appli
25	186	9.9	304	9	US-09-987-107-7	Sequence 7, Appli
26	186	9.9	304	9	US-09-987-107-8	Sequence 8, Appli
27	186	9.9	323	9	US-09-987-107-56	Sequence 56, Appli
28	186	9.9	323	9	US-09-987-107-58	Sequence 58, Appli
29	186	9.9	323	9	US-09-987-107-60	Sequence 60, Appli
30	185.5	9.9	243	9	US-09-987-107-1	Sequence 1, Appli
31	185.5	9.9	243	9	US-09-987-107-2	Sequence 2, Appli
32	185.5	9.9	261	9	US-09-987-107-52	Sequence 52, Appli
33	185.5	9.9	306	9	US-09-987-107-9	Sequence 9, Appli
34	185.5	9.9	306	9	US-09-987-107-10	Sequence 10, Appli
35	185.5	9.9	306	9	US-09-987-107-11	Sequence 11, Appli
36	185.5	9.9	324	9	US-09-987-107-62	Sequence 62, Appli
37	185.5	9.9	324	9	US-09-987-107-64	Sequence 64, Appli
38	185.5	9.9	324	9	US-09-987-107-66	Sequence 66, Appli
39	185.5	9.9	329	9	US-09-987-107-14	Sequence 14, Appli
40	185.5	9.9	336	9	US-09-987-107-44	Sequence 44, Appli
41	185.5	9.9	337	9	US-09-987-107-46	Sequence 46, Appli
42	185.5	9.9	344	9	US-09-987-107-68	Sequence 68, Appli
43	185	9.9	266	9	US-09-987-107-20	Sequence 20, Appli
44	183	9.7	264	9	US-09-987-107-25	Sequence 25, Appli
45	176	9.4	265	9	US-09-987-107-17	Sequence 17, Appli

## ALIGNMENTS

RESULT 1  
US-09-751-877-3  
Sequence 3, Application US/09751877  
Patent No. US20020142949A1  
GENERAL INFORMATION:  
APPLICANT: Yen, Frances  
APPLICANT: Denison, Blake  
APPLICANT: Bour, Barbara  
APPLICANT: Bihain, Bernard  
APPLICANT: Dumas Milne Edwards, Jean-Baptiste  
APPLICANT: Duclet, Aymeric  
APPLICANT: Bouguetel, Lydie  
APPLICANT: Ebbels-Reed, Dana  
APPLICANT: Salter-Cid, Luisa  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING NEOPLASTIC CELL GROWTH  
FILE REFERENCE: 89, US3, RFG  
CURRENT APPLICATION NUMBER: US/09/751, 877  
CURRENT FILING DATE: 2000-12-28  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: Patent, pm  
SEQ ID NO 3  
LENGTH: 366  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-751-877-3  
Query Match 100.0%, Score 1877, DB 10, Length 366;  
Best Local Similarity 100.0%, Pred. No. 1, le-124;  
Matches 366; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 MASMAVLTWALTLASAFSATOARGKFMFYSGTSGDKGRVGRDIIHQKKAAREPATLKDSL 60  
1 MASMAVLTWALTLASAFSATOARGKFMFYSGTSGDKGRVGRDIIHQKKAAREPATLKDSL 60  
61 EDDLNMMNFLEKLRPLSSSEAPRLPQDPVGNRQLOEELEVKARLPYMAEHLVGM 120  
61 EDDLNMMNFLEKLRPLSSSEAPRLPQDPVGNRQLOEELEVKARLPYMAEHLVGM 120  
121 NIEGLRQQLKPYTMDIMEQVALRVDELQRLRVGEGDTKAQLLGVDAMALLGQISRV 180  
121 NIEGLRQQLKPYTMDIMEQVALRVDELQRLRVGEGDTKAQLLGVDAMALLGQISRV 180

QY	241	ALHAKIQQNLDDLRREILSKRAFGTCTEBGACPDQOMLSEEVRORIQA*RODTYLIQIAFT	3000
Db	241	ALHARIQQNLDDLRREILSKRAFGTCTEBGACPDQOMLSEEVRORIQA*RODTYLIQIAFT	3000
QY	301	KALIDETEEVQOQLAPPPPGHSARAFAPERQOITDSQKVLSKIQARLDDIMEDIITSHLHQGH	3600
Db	301	KALIDETEEVQOQLAPPPPGHSARAFAPERQOITDSQKVLSKIQARLDDIMEDIITSHLHQGH	3600
QY	361	SHLQDP 366	
Db	361	SHLQDP 366	

```

: RESULT 2
: US-09-835-996A-2
: Sequence 2, Application US/09835996A
: Patent No. US20020142953A1
: GENERAL INFORMATION:
: APPLICANT: Ballinger, Dennis
: APPLICANT: Loeb, Debra
: APPLICANT: Montgomery, Julie
: APPLICANT: Tang, Y. Tom
: APPLICANT: Zhou, Ping
: APPLICANT: Goodrich, Ryle
: APPLICANT: Liu, Chenghua
: APPLICANT: Asundi, Vinod
: APPLICANT: Zhao, Qing
: APPLICANT: Wehrman, Tom
: APPLICANT: Drmanac, Radoje
: APPLICANT: Ren, Kejian
: APPLICANT: Qian, Xiahong
: APPLICANT: Wang, Duncui
: TITLE OF INVENTION: MATERIALS AND METHODS RELATING TO LIPID METABOLISM
: FILE REFERENCE: 28110/35915A
: CURRENT APPLICATION NUMBER: US/09/835,996A
: CURRENT FILING DATE: 2001-04-16
: PRIOR APPLICATION NUMBER: US 60/197,137
: PRIOR FILING DATE: 2000-04-14
: PRIOR APPLICATION NUMBER: US 09/714,936
: PRIOR FILING DATE: 2000-11-17
: PRIOR APPLICATION NUMBER: US 09/667,298
: PRIOR FILING DATE: 2000-09-22
: PRIOR APPLICATION NUMBER: US 09/631,451
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: US 09/598,042
: PRIOR FILING DATE: 2000-06-20
: NUMBER OF SEQ ID NOS: 45
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 2
: LENGTH: 366
: TYPE: prt
: ORGANISM: Homo sapiens
: US-09-835-996A-2

```

Query Match	Similarity	Score	1877	DB	10	length	366
Best Local	Similarity	Score	No.	1	1e-124		
Matches	366	Conservative	0	Mismatches	0	Indels	0
						Gaps	
QY	1	MASMAAVLTWALALLSAFSATQARKGFWDYFSQTSGDKGRVEQIHQOKMAREPATLNDSL	60				
DB	1	MASMAAVLTWALALLSAFSATQARKGFWDYFSQTSGDKGRVEQIHQOKMAREPATLNDSL	60				
QY	61	EODLNNMKFELEKRLPLSGSEAPRLPODPVGNRRLOLELESEYKARLOPYAAEHHELWGM	122				
DB	61	EODLNNNNKFELEKRLPLSGSEAPRLPODPVGNRRLOLELESEYKARLOPYAAEHHELWGM	120				
QY	121	NI EGLRQOLAKPYTMDLMEQVALRVOELEQOLRVVGEEDTKKQDLGVDAMALLGGLQLSRV	180				
DB	121	NI EGLRQOLAKPYTMDLMEQVALRVOELEQOLRVVGEEDTKKQDLGVDAMALLGGLQLSRV	180				

QY	181	VHHTGRFELHPHAESLVSIGHHVOELHNSVAPHAPASPARLSRCVYLSRLLTKAK	2400
DB	181	VHHTGRFELHPHAESLVSIGHHVOELHNSVAPHAPASPARLSRCVYLSRLLTKAK	2400
QY	241	ALHARIQONLDOLREBLSRAFGTGEAGAPDPOMLSEEVYRORLQAFRODTYLQIAAFT	3000
DB	241	ALHARIQONLDOLREBLSRAFGTGEAGAPDPOMLSEEVYRORLQAFRODTYLQIAAFT	3000
QY	301	RAIDDETEEVVOOQIAPPFGHSAPFPEOOTSCKVLSKLOARLDDIMEDITTHSLHOGCH	3600
DB	301	RAIDDETEEVVOOQIAPPFGHSAPFPEOOTSCKVLSKLOARLDDIMEDITTHSLHOGCH	3600
QY	361	SHLGGP 366	
DB	361	SHLGGP 366	

```

RESULT 3
US-09-835-996A-41
: Sequence 41, Application US/09835996A
: Patent No. US20020142953A1
: GENERAL INFORMATION:
: APPLICANT: Ballinger, Dennis
: APPLICANT: Loeb, Debra
: APPLICANT: Montgomery, Julie
: APPLICANT: Tang, Y. Tom
: APPLICANT: Zhou, Ping
: APPLICANT: Goodrich, Ryle
: APPLICANT: Liu, Chenghua
: APPLICANT: Asundi, Vinod
: APPLICANT: Zhao, Qing
: APPLICANT: Wehrman, Tom
: APPLICANT: Drmanac, Radoje
: APPLICANT: Ren, Feiyan
: APPLICANT: Qian, Xiaohong
: APPLICANT: Wang, Dunrui
: TITLE OF INVENTION: MATERIALS AND METHODS RELATING TO LIPID METABOLISM
: FILE REFERENCE: 28110/35915A
: CURRENT APPLICATION NUMBER: US/09/835,996A
: PRIOR FILING DATE: 2001-04-16
: PRIOR APPLICATION NUMBER: US 60/197,137
: PRIOR FILING DATE: 2000-04-14
: PRIOR APPLICATION NUMBER: US 09/714,936
: PRIOR FILING DATE: 2000-11-17
: PRIOR APPLICATION NUMBER: US 09/667,298
: PRIOR FILING DATE: 2000-09-22
: PRIOR APPLICATION NUMBER: US 09/631,451
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: US 09/598,042
: PRIOR FILING DATE: 2000-06-20
: NUMBER OF SEQ ID NOS: 45
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 41
: LENGTH: 400
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-835-996A-41

```

	Query Match	Similarity	Score	DB	Length
Best Local	100.0%	100.0%	1877	12	400
Matches	366	Conservative	0	Mismatches	0
				Indels	0
				Gaps	0
QY	1	MASMAAVLTWALIALISAFSAQTQARKGFWDYFSQTSDDKGRVEDIHOOKKAREPATITKDSL	60		
DB	35	MASMAAVLTWALIALISAFSAQTQARKGFWDYFSQTSDDKGRVEDIHOOKKAREPATITKDSL	94		
QY	61	EDDLNNKFKLEKRLPLSSSEAPRLPQDPVGMKRRQGELEEVKARIQPYMAEAHELVGM	120		
DB	95	EDDLNNKFKLEKRLPLSSSEAPRLPQDPVGMKRRQGELEEVKARIQPYMAEAHELVGM	154		
QY	121	NIEGLRQOLKPTTMDIMEQVALRVDELQBLRVYGEDTQAOLLGVDYEMALLQIQISRY	180		

Db 155 NLEGLRQOLKPYTMDLMEQVALRVOELQVLRVVGEDTKAQLGCVDEANALLQGLQSRV 214

QY 181 VHTTGRKELFHPYAESLVSGIGHVOELHRSVAHPASPAPSLRCVQVLSRKLTLKAK 240

Db 215 VHTTGRKELFHPYAESLVSGIGHVOELHRSVAHPASPAPSLRCVQVLSRKLTLKAK 274

QY 241 ALHARLQONLDOLREELSRFAGTGTGEGAGPPDQMLSEEVRORLQAFRODTYLOIAAFT 300

Db 275 ALHARLQONLDOLREELSRFAGTGTGEGAGPPDQMLSEEVRORLQAFRODTYLOIAAFT 334

QY 301 RAIDQETEEVQOQLAPPFGHSAFAPEFQOTDSGKVLSTKLQARLDLMDITTHSLHDOGH 360

Db 335 RAIDQETEEVQOQLAPPFGHSAFAPEFQOTDSGKVLSTKLQARLDLMDITTHSLHDOGH 394

QY 361 SHLQDP 366

Db 395 SHLQDP 400

## RESULT 4

US-09-800-729-212

; Sequence 212, Application US/09800729

; Patent No. US20020068319A1

; GENERAL INFORMATION:

; APPLICANT: Ni et al.

; TITLE OF INVENTION: 32 Human secreted proteins

; FILE REFERENCE: P2044P1

; CURRENT APPLICATION NUMBER: US/09/800,729

; PRIOR FILING DATE: 2001-03-08

; PRIOR APPLICATION NUMBER: PCT/US00/26013

; PRIOR FILING DATE: 2000-09-22

; PRIOR APPLICATION NUMBER: 60/155,709

; PRIOR FILING DATE: 1999-09-24

; NUMBER OF SEQ ID NOS: 217

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 212

; LENGTH: 366

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-800-729-212

Query Match 99.7%; Score 1871; DB 10; Length 366;

Best Local Similarity 99.7%; Pred. No. 2.8e-124;

Matches 365; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAAVAITLWALLLSAFSATOARKGFWDYFSQTSQDKGRVEQIHQOKMAREPATLKDLSL 60

Db 1 MAAVAITLWALLLSAFSATOARKGFWDYFSQTSQDKGRVEQIHQOKMAREPATLKDLSL 60

QY 61 EODLNNNKFLFKLRPLSGSPAPRLPDPVGMRRLOQLBELLEVYKARIQPYMAFAHELVGM 120

Db 61 EODLNNNKFLFKLRPLSGSPAPRLPDPVGMRRLOQLBELLEVYKARIQPYMAFAHELVGM 120

QY 121 NLEGLRQOLKPYTMDLMEQVALRVOELQVLRVVGEDTKAQLGCVDEANALLQGLQSRV 180

Db 121 NLEGLRQOLKPYTMDLMEQVALRVOELQVLRVVGEDTKAQLGCVDEANALLQGLQSRV 180

QY 181 VHTTGRKELFHPYAESLVSGIGHVOELHRSVAHPASPAPSLRCVQVLSRKLTLKAK 240

Db 181 VHTTGRKELFHPYAESLVSGIGHVOELHRSVAHPASPAPSLRCVQVLSRKLTLKAK 240

QY 241 ALHARLQONLDOLREELSRFAGTGTGEGAGPPDQMLSEEVRORLQAFRODTYLOIAAFT 300

Db 241 ALHARLQONLDOLREELSRFAGTGTGEGAGPPDQMLSEEVRORLQAFRODTYLOIAAFT 300

QY 301 RAIDQETEEVQOQLAPPFGHSAFAPEFQOTDSGKVLSTKLQARLDLMDITTHSLHDOGH 360

Db 301 RAIDQETEEVQOQLAPPFGHSAFAPEFQOTDSGKVLSTKLQARLDLMDITTHSLHDOGH 360

QY 361 SHLQDP 366

Db 361 SHLQDP 366

## RESULT 5

US-09-800-729-111

; Sequence 111, Application US/09800729

; Patent No. US20020068319A1

; GENERAL INFORMATION:

; APPLICANT: Ni et al.

; TITLE OF INVENTION: 32 Human secreted proteins

; FILE REFERENCE: P2044P1

; CURRENT APPLICATION NUMBER: US/09/800,729

; PRIOR FILING DATE: 2001-03-08

; PRIOR APPLICATION NUMBER: PCT/US00/26013

; PRIOR FILING DATE: 2000-09-22

; PRIOR APPLICATION NUMBER: 60/155,709

; PRIOR FILING DATE: 1999-09-24

; NUMBER OF SEQ ID NOS: 217

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 111

; LENGTH: 363

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-800-729-111

Query Match 99.0%; Score 1858; DB 10; Length 363;

Best Local Similarity 99.7%; Pred. No. 2.3e-123;

Matches 362; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 MAAVAITLWALLLSAFSATOARKGFWDYFSQTSQDKGRVEQIHQOKMAREPATLKDLSL 63

Db 1 MAAVAITLWALLLSAFSATOARKGFWDYFSQTSQDKGRVEQIHQOKMAREPATLKDLSL 60

QY 64 LNNNNKFLFKLRPLSGSPAPRLPDPVGMRRLOQLBELLEVYKARIQPYMAFAHELVGM 123

Db 61 LNNNNKFLFKLRPLSGSPAPRLPDPVGMRRLOQLBELLEVYKARIQPYMAFAHELVGM 120

QY 124 GLRQOLKPYTMDLMEQVALRVOELQVLRVVGEDTKAQLGCVDEANALLQGLQSRV 183

Db 121 GLRQOLKPYTMDLMEQVALRVOELQVLRVVGEDTKAQLGCVDEANALLQGLQSRV 180

QY 184 TGRKELFHPYAESLVSGIGHVOELHRSVAHPASPAPSLRCVQVLSRKLTLKAK 243

Db 181 TGRKELFHPYAESLVSGIGHVOELHRSVAHPASPAPSLRCVQVLSRKLTLKAK 240

QY 244 ARIQONLDOLREELSRFAGTGTGEGAGPPDQMLSEEVRORLQAFRODTYLOIAAFT 303

Db 241 ARIQONLDOLREELSRFAGTGTGEGAGPPDQMLSEEVRORLQAFRODTYLOIAAFT 300

QY 304 DOETEEVQOQLAPPFGHSAFAPEFQOTDSGKVLSTKLQARLDLMDITTHSLHDOGH 363

Db 301 DOETEEVQOQLAPPFGHSAFAPEFQOTDSGKVLSTKLQARLDLMDITTHSLHDOGH 360

QY 364 GDP 366

Db 361 GDP 363

## RESULT 6

US-09-987-107-37

; Sequence 37, Application US/0987107

; Patent No. US20020156007A1

; GENERAL INFORMATION:

; APPLICANT: GRAVERSEN, Jonas

; TITLE OF INVENTION: APOLOPROTEIN ANALOGUES

; FILE REFERENCE: GRAVERSENIA

; CURRENT APPLICATION NUMBER: US/09/987,107

; PRIOR FILING DATE: 2001-11-13

; PRIOR APPLICATION NUMBER: US 60/264,022

; PRIOR FILING DATE: 2001-01-26

; PRIOR APPLICATION NUMBER: DK PA2001 00057

; PRIOR FILING DATE: 2001-01-15

; PRIOR APPLICATION NUMBER: DK PA2000 01682

; PRIOR FILING DATE: 2000-11-10



```

GENERAL INFORMATION:
APPLICANT: Ni et al.
TITLE OF INVENTION: 32 Human secreted proteins
FILE REFERENCE: P2044P1
CURRENT APPLICATION NUMBER: US/09/800,729
CURRENT FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: PCT/US00/26013
PRIOR FILING DATE: 2000-09-22
PRIOR APPLICATION NUMBER: 60/155,709
PRIOR FILING DATE: 1999-09-24
NUMBER OF SEQ ID NOS: 217
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 207
LENGTH: 396
TYPE: PRT
ORGANISM: Homo sapiens
US-09-800-729-207

```

```

Query Match      17.5%; Score 328; DB 10; Length 396;
Best Local Similarity 28.4%; Pred. No. 4.5e-16;
Matches 96; Conservative 72; Mismatches 152; Indels 18; Gaps 6;

```

```

QY 5 AAVLTALALLSLA-----FSATQARKGFMDYFSQTSQD-KGRVQIHQOKMARE-PATLKD 58
    ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| |||
DB 5 AAVLTALAVAGARAEVSADQAVTVMWYFSSLSNNAKEAVEHLQKSEITQQLNALPQD 64
QY 59 SLEQDNLNMNKKFLEKRLPLSGSEAPRLPDPVGMKROLOEELFEVKARLQPYMAEAHEL 118
    ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| |||
DB 65 KLGSEVNTYAGDLOKLVFPFTELHERLAKNSEKKEIGLEELERALLPHANEVSQKI 124
QY 119 GWNLEGIROOLKPYTMDLMEQVALRVQEIQOQLRVGCEDTKAQLGCVDAAMALIQ----- 174
    ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| |||
DB 125 GDNIRELOQRLEPYADQLRQVNTQAEQLRRQDPLAQRERVLRENADSLQASLRPHAD 184
QY 175 GLOSRYVHHGTRFKELEHPYAESLSVSGIGHVOELHRSVAPHPASPARLSRCVQVLSRK 234
    ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| |||
DB 185 ELAKKIDQNVDELKGRITPYADEFKVKIIDQVEELKRSLLAPYADYQKILNHQILGILTFQ 244
QY 235 LTIKAKALHARIQONLDQRLRELSRAFACT-----GTEEGAGDPQMLSEEVKQRIQAFR 289
    ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| |||
DB 245 MKKNAEELKARISASAEELRQRLAPLAEDVRGNLKGTEGLQSLAELGHIIDQVEEFR 304
QY 290 QDTYLIQIAATRAIDQTEVEVQOQLAP---PPGHSAP 324
    ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| |||
DB 305 RVEPYGENFNKALVQDMQEQLRKILGPHAGDVGHLSF 342

```

## RESULT 10

```

US-09-987-107-33
Sequence 33; Application US/09987107
Patent No. US20020156007A1

```

```

GENERAL INFORMATION:
APPLICANT: GRAVERSEN, Jonas
APPLICANT: MOESTRUP, Soren
TITLE OF INVENTION: APOLIPROTEINS ANALOGUES
FILE REFERENCE: GRAVERSENIA
CURRENT APPLICATION NUMBER: US/09/987,107
CURRENT FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: US 60/264,022
PRIOR FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: DK PA2001 00057
PRIOR FILING DATE: 2001-01-15
PRIOR APPLICATION NUMBER: DK PA2000 01682
PRIOR FILING DATE: 2000-11-10
NUMBER OF SEQ ID NOS: 91
SOFTWARE: PatentIn version 3.1
SEQ ID NO 33
LENGTH: 396
TYPE: PRT
ORGANISM: Homo sapiens
US-09-987-107-33

```

```

Query Match      17.4%; Score 327; DB 9; Length 396;

```

```

Best Local Similarity 28.4%; Pred. No. 5.3e-16;
Matches 96; Conservative 71; Mismatches 153; Indels 18; Gaps 6;

```

```

QY 5 AAVLTALALLSLA-----FSATQARKGFMDYFSQTSQD-KGRVQIHQOKMARE-PATLKD 58
    ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| |||
DB 5 AAVLTALAVAGARAEVSADQAVTVMWYFSSLSNNAKEAVEHLQKSEITQQLNALPQD 64
QY 59 SLEQDNLNMNKKFLEKRLPLSGSEAPRLPDPVGMKROLOEELFEVKARLQPYMAEAHEL 118
    ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| |||
DB 65 KLGSEVNTYAGDLOKLVFPFTELHERLAKNSEKKEIGLEELERALLPHANEVSQKI 124
QY 119 GWNLEGIROOLKPYTMDLMEQVALRVQEIQOQLRVGCEDTKAQLGCVDAAMALIQ----- 174
    ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| |||
DB 125 GDNIRELOQRLEPYADQLRQVNTQAEQLRRQDPLAQRERVLRENADSLQASLRPHAD 184
QY 175 GLOSRYVHHGTRFKELEHPYAESLSVSGIGHVOELHRSVAPHPASPARLSRCVQVLSRK 234
    ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| |||
DB 185 ELAKKIDQNVDELKGRITPYADEFKVKIIDQVEELKRSLLAPYADYQKILNHQILGILTFQ 244
QY 235 LTIKAKALHARIQONLDQRLRELSRAFACT-----GTEEGAGDPQMLSEEVKQRIQAFR 289
    ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| |||
DB 245 MKKNAEELKARISASAEELRQRLAPLAEDVRGNLKGTEGLQSLAELGHIIDQVEEFR 304
QY 290 QDTYLIQIAATRAIDQTEVEVQOQLAP---PPGHSAP 324
    ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| |||
DB 305 RVEPYGENFNKALVQDMQEQLRKILGPHAGDVGHLSF 342

```

## RESULT 11

```

US-09-987-107-36
Sequence 36; Application US/09987107
Patent No. US20020156007A1

```

```

GENERAL INFORMATION:
APPLICANT: GRAVERSEN, Jonas
APPLICANT: MOESTRUP, Soren
TITLE OF INVENTION: APOLIPROTEINS ANALOGUES
FILE REFERENCE: GRAVERSENIA
CURRENT APPLICATION NUMBER: US/09/987,107
CURRENT FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: US 60/264,022
PRIOR FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: DK PA2001 00057
PRIOR FILING DATE: 2001-01-15
PRIOR APPLICATION NUMBER: DK PA2000 01682
PRIOR FILING DATE: 2000-11-10
NUMBER OF SEQ ID NOS: 91
SOFTWARE: PatentIn version 3.1
SEQ ID NO 36
LENGTH: 401
TYPE: PRT
ORGANISM: Papio anubis
US-09-987-107-36

```

```

Query Match      16.8%; Score 315; DB 9; Length 401;
Best Local Similarity 28.1%; Pred. No. 3.7e-15;
Matches 92; Conservative 70; Mismatches 137; Indels 28; Gaps 8;

```

```

QY 19 SATQARKGFMDYFSQTSQD-KGRVQIHQOKMARE-PATLKDSELDLNMNKKFLEKLR 76
    ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| |||
DB 7 SADQAVTVMWYFSSLSNNAKEAVEHLQKSEITQQLNALPQDQVNTYAGDLOKLV 66
QY 77 LSGSEAPRLPDPVGMKROLOEELFEVKARLQPYMAEAHELQVNMNLEGLRQOLKPYTMDL 136
    ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| |||
DB 67 FATELHERLAKDSKKLEELRKELEERALLPHANEVSQKIENRELOQRLEPTDQL 126
QY 137 MEQVALRVQEIQOQLRVGCEDTKAQLGCVDAAMALIQ-----GLOSRYVHHGTRFKELE 192
    ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| |||
DB 127 RTQVNTQTEQLRQRLTPYADQRMERVLRENADSLQTSLRPHADQKAKIDQNVDELKGRILT 186
QY 193 PYAESLSVSGIGHVOELHRSVAPHPASPARLSRCVQVLSRKILTIKAKALHARIQONLDQ 252
    ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| |||
DB 187 PYADEFKVKIIDQVEELKRSLLAPYADYQKILNHQILGILGILAPFQKNAEELKARISASAE 246

```

OY 253 LREE--LSRAFACT--GTEGAGPDPOMLS-----EEVORLQARODTYIOIAFT 300  
 DB 247 LKORLAPIAEDMRGLKNTCELOKSLAELOCHLDHNVFEERLKEVPEGEN-----FN 299  
 OY 301 RAIDQTEEVQOOLAP---PPGHSAP 324  
 DB 300 KALVQOEMQLRQKLPAGHAGDVEGHSF 326

# RESULT 12

US-09-987-107-38  
 ; Sequence 38, Application US/09987107  
 ; Patent No. US20020156007A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: GRAVERSEN, Jonas  
 ; APPLICANT: MOESTRUP, Soren  
 ; TITLE OF INVENTION: APOLIPOPROTEINS ANALOGUES  
 ; FILE REFERENCE: GRAVERSENIA  
 ; CURRENT APPLICATION NUMBER: US/09/987,107  
 ; CURRENT FILING DATE: 2001-11-13  
 ; PRIOR APPLICATION NUMBER: US 60/264,022  
 ; PRIOR FILING DATE: 2001-01-26  
 ; PRIOR APPLICATION NUMBER: DK PA2001 00057  
 ; PRIOR FILING DATE: 2001-01-15  
 ; PRIOR APPLICATION NUMBER: DK PA2000 01682  
 ; PRIOR FILING DATE: 2000-11-10  
 ; NUMBER OF SEQ ID NOS: 91  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO: 38  
 ; LENGTH: 391  
 ; TYPE: PRT  
 ; ORGANISM: Rattus norvegicus  
 ; US-09-987-107-38

Query Match 16.6%; Score 311.5; DB 9; Length 391;  
 Best Local Similarity 24.8%; Pred. No. 6.4e-15;  
 Matches 95; Conservative 84; Mismatches 151; Indels 53; Gaps 9;

OY 5 AAVLTWALALISAFA--TQARKCWDYFSQTSQD-KGRAVEQIHQOKMAREPATLKOS 59  
 DB 5 AAVLTWALALISAFA--TQARKCWDYFSQTSQD-KGRAVEQIHQOKMAREPATLKOS 61  
 OY 60 LEODLNMMKFE---KLRLPSSEAPRLPDQPVGMRRQLOEELVEYKARLOPYMAEAM 115  
 DB 62 FQDKLGNITTYADDLQNKLVPRAYQSLGHLTKETERVREIQELEDLANNMHPANKXS 121  
 OY 116 ELVGMNLKQLRQOLKTYTMDLMROYALRYQELQEQLRVVGCHTKAQLGCVDPAMALLQG 175  
 DB 122 QMGEDVVOKLOEHLREYATDLOAIQAOTQDMKROLTPYIQHQTIDQNV-----EN 174  
 OY 176 LOSRVVHHTGCRKELPH-----PYARSIVSGICRHVOELHRSVAPHPAPASPARI 224  
 DB 175 LOSRVVHHTGCRKELPH-----PYARSIVSGICRHVOELHRSVAPHPAPASPARI 224  
 OY 225 SRVVOYLKRKLTLKAKALHARIQOINLQLELSRAFACT--GTEGAGPDPOMLSE 279  
 DB 235 NHQMEELAFQOMKKNABELQTKVSTINIDQOKNLAPLVEDVQSKLKGNTGCIQKSLIEDLAK 294  
 OY 280 EYVQRQIARQODTYIOIAFTRAIDQETREYVQOQLAPPPGHSAPAPPEQOQIDSGV--- 336  
 DB 295 QLDQOYVEFRRAVEPLGDFNNALVQOMKEFRQQLG-----SDSGGVESH 339  
 OY 337 LSKLQARLDLMDLDTTHSLHDCG 359  
 DB 340 LSLLEKNLREKVSFFMSTLQKKG 362

RESULT 13  
 US-09-800-729-208  
 ; Sequence 208, Application US/09800729  
 ; Patent No. US20020068319A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: NI et al.

; TITLE OF INVENTION: 32 Human secreted proteins  
 ; FILE REFERENCE: P2044PI  
 ; CURRENT APPLICATION NUMBER: US/09/800,729  
 ; CURRENT FILING DATE: 2001-03-08  
 ; PRIOR APPLICATION NUMBER: PCT/US00/26013  
 ; PRIOR FILING DATE: 2000-09-22  
 ; PRIOR APPLICATION NUMBER: 60/155,709  
 ; PRIOR FILING DATE: 1999-09-24  
 ; NUMBER OF SEQ ID NOS: 217  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO: 208  
 ; LENGTH: 391  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-09-800-729-208

Query Match 15.9%; Score 298.5; DB 10; Length 391;  
 Best Local Similarity 24.4%; Pred. No. 5.2e-14;  
 Matches 91; Conservative 85; Mismatches 164; Indels 33; Gaps 7;

OY 5 AAVLTWALALISA--FSATQARKGFWDFYFSQTSQD-KGRAVEQIHQOKMAREPATLK-KD 58  
 DB 5 AAVLTWALALISA--FSATQARKGFWDFYFSQTSQD-KGRAVEQIHQOKMAREPATLK-KD 64  
 OY 59 SLEODLNMMKFEKLRPLSSSEAPRLPDQPVGMRRQLOEELVEYKARLOPYMAEAM 118  
 DB 65 KLGDASTYADGVHNLVFPVVOVLSGLAKETERYKEELKKLELEDLRQMMHPANKVOTPF 124  
 OY 119 GWNLEGLRQOLKPYTMDIMEQVALRYQELQEQLRVVGCHTKAQLGCVDPAMALLQG 174  
 DB 125 GEMOKLOEHLKPYAVDLOQDINTQOEMKLOLTPYIORMOTTIKENDNLHTSMPLAT 184  
 OY 175 GLQSRVHHTGCRKELPHRYAASIVSGICRHVOELHRSVAPHPAPASPARISKV 234  
 DB 185 NLKDFNRMMEELKHLTPRANLTKATIDONLEDIRSLAPLTVCQVKLNHMEGLAFQ 244  
 OY 235 LTLKAKALHARIQOINLQLELSRAFACT--GTEGAGPDPOMLSEEVORLQAR 289  
 DB 245 MKKNABELQTKSAKIDQOKNLAPLVEDVQSKYKNGTEGLOKSLIEDLNRLEQVEEFR 304  
 OY 290 QDTYIOIAFTRAIDQETREYVQOQLAPPPGHSAPAPPEQOQIDSGV---LSKLQARLD 346  
 DB 305 RIVERMGEENKALVQOQLEQFRQQLG-----NSGEVSHLSFLEKSLRE 349  
 OY 347 LMEDLTHSLHDCG 359  
 DB 350 KVNSTPMSTLEKKG 362

RESULT 14  
 US-09-987-107-35  
 ; Sequence 35, Application US/09987107  
 ; Patent No. US20020156007A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: GRAVERSEN, Jonas  
 ; APPLICANT: MOESTRUP, Soren  
 ; TITLE OF INVENTION: APOLIPOPROTEINS ANALOGUES  
 ; FILE REFERENCE: GRAVERSENIA  
 ; CURRENT APPLICATION NUMBER: US/09/987,107  
 ; CURRENT FILING DATE: 2001-11-13  
 ; PRIOR APPLICATION NUMBER: US 60/264,022  
 ; PRIOR FILING DATE: 2001-01-26  
 ; PRIOR APPLICATION NUMBER: DK PA2001 00057  
 ; PRIOR FILING DATE: 2001-01-15  
 ; PRIOR APPLICATION NUMBER: DK PA2000 01682  
 ; PRIOR FILING DATE: 2000-11-10  
 ; NUMBER OF SEQ ID NOS: 91  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO: 35  
 ; LENGTH: 395  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 ; US-09-987-107-35



Query Match 15.9%: Score 298.5; DB 9; Length 395;

Best Local Similarity 24.4%: Pred. No. 5.2e-14; Mismatches 164; Indels 33; Gaps 7;

Matches 91; Conservative 85; Mismatches 164; Indels 33; Gaps 7;

QY 5 AAVLTWALALSA-----FSATQARKGFMDYFSQTSGD-KGRVEQIHQOKMAREPATL-KD 58

DB 5 AAVLTWALALSA-----FSATQARKGFMDYFSQTSGD-KGRVEQIHQOKMAREPATL-KD 58

QY 59 SLFQDLNMMKFLKLRPLSGSEAPRLPQDPVGMRRQLOEELEVKARIQPMAEAHELY 118

DB 65 KLGDASTYADGVHNNKLVFVVOQLSGHLAKETEKKELEEDLRDMPIHANKVTOTF 124

QY 119 GWNLEGLRQOLKPYTMDMEQVALRVQELQVGVGDTKQALLGYD---EAMALLQ 174

DB 125 GENMCKLOEHLKPYAVDLODQINTOTQEMKIQITPYIQMOTTIKENVDNLHTSMPLAT 184

QY 175 GLQSRVHHHTGFKELFHPYAESLVSGIRHVOELHRSVAPAPASPARLSRCVYLSRK 234

DB 185 NLKDFNRNMEELKGLHLPANLEKATIDQNLDELRLSLAPLVGVQERLNMHOMGLAFQ 244

QY 235 LTLKAKALHARIOQLDOLRELSRAFGT-----GTEGAGDPDQMLSHVROHLOAFR 289

DB 245 MKKNAEELQTKVSAKIDQLOKLAPLVEDVQSKVKGNTBGLQSLLEDLRQLEQVVEFR 304

QY 290 QDTYLIQIAFTRAIDQETEEVQOQLAPPPGHSAPAFEFQOTDSKV---LSKLOARLDD 346

DB 305 RTVEPMGUMFNKALVQOLEQFROQIGP-----NSGEVESHLSFILEKSLRE 349

QY 347 LMEDITHSLHDG 359

DB 350 KVNSEPMSTLEKKG 362

RESULT 15

US-09-987-107-15

Sequence 15, Application US/09987107

Patent No. US20020156007A1

GENERAL INFORMATION:

APPLICANT: GRAVERSEN, Jonas

APPLICANT: MOESTRUP, Soren

TITLE OF INVENTION: APOLIPOPROTEINS ANALOGUES

FILE REFERENCE: GRAVERSENIA

CURRENT APPLICATION NUMBER: US/09/987,107

PRIOR FILING DATE: 2001-11-13

PRIOR APPLICATION NUMBER: US 60/264,022

PRIOR FILING DATE: 2001-01-26

PRIOR APPLICATION NUMBER: DK PA2001 00057

PRIOR FILING DATE: 2001-01-15

PRIOR APPLICATION NUMBER: DK PA2000 01682

PRIOR FILING DATE: 2000-11-10

NUMBER OF SEQ ID NOS: 91

SOFTWARE: PatentIn version 3.1

SEQ ID NO 15

LENGTH: 267

TYPE: PRT

ORGANISM: Homo sapiens

US-09-987-107-15

Query Match 10.7%: Score 200; DB 9; Length 267;

Best Local Similarity 22.2%: Pred. No. 2.6e-07;

Matches 74; Conservative 61; Mismatches 106; Indels 92; Gaps 10;

DB 119 AKVQPYLDQFQKKMOEMELRYRQKVEPLRAELQEGARQKTLHELQKLSPLGEMR----- 173

QY 165 GVDEAMALLQGLQSRVHHHTGFKELFHPYAESLVSGIRHVOELHRSVAPAPASPARL 224

DB 174 --DKARA-----HVDALRTHLAPYSDLRQL 198

QY 225 SRCVQVLSRKLTLKAKALHARIOQLDOLRELSRAFGTGEAGAPDPQMLSEVROR 284

DB 199 AARLEALKENGARLAELYHAKATEHLSTLSEKAKPAL-----EDLRQG 241

QY 285 ---LQAFRQDTYLIQIAFTRAIDQETEEVQO 313

DB 242 LLPVLESEK-----VSFLSALLEETKLNQ 267

Search completed: January 23, 2003, 17:48:14

Job time: 41 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: January 23, 2003, 17:41:21 : Search time 38 Seconds  
(without alignments)  
925.926 Million cell updates/sec

Title: US-09-842-364-3

Perfect score: 1877

Sequence: 1 MASMAVLTWMLALLSASFSA.....LMEDITHSLHDGSHLGD 366

Scoring table:

BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR73:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	333	17.7	429	2	S29565
2	328	17.5	396	1	IPHU44
3	319.5	17.0	391	1	IPRTA4
4	315	16.8	401	2	A47141
5	298.5	15.9	391	2	B40892
6	298.5	15.9	395	2	A40892
7	289.5	15.4	399	2	C40892
8	288	15.3	394	2	A25281
9	200	10.7	267	1	IPHU41
10	193	10.3	267	1	JS0079
11	187	10.0	267	1	A26529
12	185	9.9	266	1	LPRB12
13	183	9.7	266	1	LPRB12
14	180	9.6	265	1	LPRB18
15	178	9.5	265	2	AC5858
16	177	9.4	329	2	AC5856
17	175.5	9.4	266	1	LPG41
18	170.5	9.1	241	2	A24998
19	170	9.1	265	2	JT0672
20	168	9.0	265	2	A46018
21	168	9.0	317	2	A28792
22	168	9.0	317	2	S03185
23	165	8.8	317	2	S33450
24	163.5	8.7	1162	2	D83454
25	161	8.6	264	2	S22420
26	158.5	8.4	246	2	A61448
27	158.5	8.4	264	2	S31394
28	155	8.3	264	2	JC5456
29	155	8.3	311	2	JU0036

30	154.5	8.2	262	2	JC1237	apolipoprotein A-I
31	154.5	8.2	317	1	LPRE	apolipoprotein E p
32	153	8.2	317	1	LPHOE	apolipoprotein E p
33	152.5	8.1	291	2	C60940	apolipoprotein E p
34	152.5	8.1	298	2	S12635	apolipoprotein E p
35	152.5	8.1	986	2	T10754	cis-goldi matrix p
36	151.5	8.1	1956	2	T16416	hypothetical prote
37	151	8.0	259	2	A24700	apolipoprotein A-I
38	150.5	8.0	316	2	S26478	apolipoprotein E -
39	150.5	8.0	1034	2	T32297	hypothetical prote
40	149	7.9	311	2	A45951	apolipoprotein E p
41	147.5	7.9	231	2	J00704	apolipoprotein A-I
42	147.5	7.9	316	2	I45996	apolipoprotein E -
43	147.5	7.9	1248	2	AH2637	conserved hypothet
44	146	7.8	316	2	JC6549	apolipoprotein E p
45	144.5	7.7	164	2	S21830	apolipoprotein A-I

## ALIGNMENTS

### RESULT 1

S29565

apolipoprotein A-IV - crab-eating macaque

C:Species: Macaca fascicularis (crab-eating macaque)

C:Date: 06-Jan-1995 #sequence-revision 06-Jan-1995 #ext\_change 13-Aug-1999

C:Accession: S30195; S29565

R:Osada, J., Pocovi, M., Nicolosi, R.J., Schaefer, E.J., Ordovas, J.M.

Biochim. Biophys. Acta 1172, 335-339, 1993

A:Title: Nucleotide sequences of the Macaca fascicularis apolipoprotein C-III and A-

A:Reference number: S30195; M01D:93192330; PMID:8448212

A:Accession: S30195

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1429 <OSA>

A:Cross-references: EMBL:X68361; NID:q38050; PIDN:CAA48421.1; PID:q38051

C:Genetics:

A:Introns: 17/1: 59/2

C:Superfamily: apolipoprotein A-I

Query Match 17.7% Score 333; DB 2; Length 429;

Best Local Similarity 28.7% Pred. No. 4.5e-13;

Matches 99; Conservative 72; Mismatches 142; Indels 32; Gaps 9;

OY	5	AAVLTWMLALISA---FSATQARKGFWDYFSQTSQD-KGRVEDIHOQKARE-PATLKD	58
DB	5	AAVLTWMLALISA---FSATQARKGFWDYFSQTSQD-KGRVEDIHOQKARE-PATLKD	64
OY	59	SLEODLNMMKFLKRLPLSGSEAPRLPOPVGMRROLOELEVYKARLOPYMAEHELV	118
DB	65	KIDSVNTYAGDLOKLVFPFATLHEKRLAKDSEKLEKELEVERKARLPHANEVSOKI	124
OY	119	GMNLEGLRQOLKPYTMDLQVVALRVOELOEOLRVVEEDPKAOLGVDAMALIQ----	174
DB	125	GENVRELQGLRLEPYTDQLRQVNTQTEQLRQRLTPYQRMRENVLRNADLSLRHAD	184
OY	175	GIQSRVNNHTGRKPELPHYAESLVSGTGRHVOELHRSVAPHPAPASPARLSKCVYLRSK	234
DB	185	QLAKADQNVNELKERLTLPYADEPKVYIDQTVELRSLAPYQDADEKLNHLEGLAFQ	244
OY	235	LTGKAKALHARIOONDLOREE---LSRAFGT--GTGEGAGDPPQMLS-----EVR	282
DB	245	MKNNAEELKARISASAEELQRLAPLAEDMKKCNLRGTBGLQSLAEGLDRLHYEEFR	304
OY	283	QRQAEFRQDTYLOIAETRAIDETEVEVQQLAP---PPGHSAP	324
DB	305	LREVEYGEN-----FNKALVOOMEQLRQKLGPHADGVGHLSF	342

### RESULT 2

LPHU44

apolipoprotein A-IV precursor [validated] - human

N:Alternate names: apoA-IV

C:Species: *Hom. sapiens* (man)  
C:Date: 30-Sep-1987 #sequence revision 30-Sep-1987 #text change 08-Dec-2000  
C:Accession: A04137; A94059; A24444; A29330; A26280; 137177; C54223; A61203; A26481; S07  
R:Karathanasis, S.K.; Oelgelgen, F.; Haddad, I.A.; Antonarakis, S.E.  
Proc. Natl. Acad. Sci. U.S.A. 83, 8457-8461, 1986  
A:Title: Structure, evolution, and polymorphisms of the human apolipoprotein A4 gene (A4)  
A:Reference number: A94137; MUID:87044474; PMID:3095836  
A:Accession: A94137  
A:Molecule type: DNA  
A:Residues: 1-396 <KAR1>  
A:Cross-references: GB:M14642; NID:q178760; PIDN:AAA51745.1; PID:q178761  
R:Karathanasis, S.K.  
Proc. Natl. Acad. Sci. U.S.A. 82, 6374-6378, 1985  
A:Title: Apolipoprotein multigene family: tandem organization of human apolipoprotein A4  
A:Reference number: A94059; MUID:86016704; PMID:3931073  
A:Accession: A94059  
A:Molecule type: mRNA  
A:Residues: 135-378 <KAR2>  
A:Cross-references: GB:M10373; NID:g563319; PIDN:AAB59516.1; PID:g563320  
R:Karathanasis, S.K.; Yunis, I.  
Biochemistry 25, 3962-3970, 1986  
A:Title: Structure, evolution, and tissue-specific synthesis of human apolipoprotein A1  
A:Reference number: A24449; MUID:86296629; PMID:3755616  
A:Accession: A24449  
A:Molecule type: mRNA  
A:Residues: 1-396 <KAR3>  
A:Cross-references: GB:M13654; NID:q178758; PIDN:AAA51744.1; PID:q178759  
R:Gordon, J.I.; Bisgaier, C.L.; Sims, H.F.; Sachdev, O.P.; Gluckman, R.M.; Strauss, A.W.  
J. Biol. Chem. 259, 4684-4774, 1984  
A:Title: Biosynthesis of human preapolipoprotein A-IV.  
A:Contents: annotation; signal sequence cleavage site  
R:Shounbady, N.A.; Walker, D.W.; Palk, Y.K.; Boguski, M.S.; Freeman, M.; Gordon, J.I.  
J. Biol. Chem. 262, 7973-7981, 1987  
A:Title: Structure and expression of the human apolipoprotein A-IV gene.  
A:Reference number: A29330; MUID:87250378; PMID:3036793  
A:Accession: A29330  
A:Molecule type: DNA  
A:Residues: 1-157, 'T', 159, 'Y', 161-278, 'R', 280-326, 'T', 328-379, 'H', 381-396 <EL5>  
A:Cross-references: GB:J02758; NID:q178756; PIDN:AAA96731.1; PID:q178757  
R:Shounbady, N.A.; Walker, D.W.; Boguski, M.S.; Gordon, J.I.; Taylor, J.M.  
J. Biol. Chem. 261, 1998-2002, 1986  
A:Title: The nucleotide and derived amino acid sequence of human apolipoprotein A-IV mRNA  
A:Reference number: A26280; MUID:86111885; PMID:3080432  
A:Accession: A26280  
A:Molecule type: mRNA  
A:Residues: 21-157, 'T', 159, 'Y', 161-278, 'R', 280-326, 'T', 328-379, 'H', 381-396 <EL2>  
A:Cross-references: GB:M14566; NID:g178778; PIDN:AAA51748.1; PID:q178779  
R:Yang, C.Y.; Gu, Z.W.; Chong, L.S.; Xiong, W.J.; Rosseneu, M.; Yang, H.X.; Lee, B.R.; G  
Biochim. Biophys. Acta 1002, 231-237, 1989  
A:Title: The primary structure of human apolipoprotein A-IV.  
A:Reference number: 137177; MUID:89194198; PMID:2930771  
A:Accession: 137177  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-157, 'T', 159, 'Y', 161-278, 'R', 280-336 <YAN1>  
A:Cross-references: EMBL:X13629; NID:q28761; PIDN:CAA1955.1; PID:q28762  
A:Note: Submitted to the EMBL Data Library, January 1989  
R:Kunitake, S.T.; Carilli, C.T.; Lau, K.; Procter, A.A.; Naya-Vigene, J.; Kane, J.P.  
Biochemistry 33, 1988-1993, 1994  
A:Title: Identification of proteins associated with apolipoprotein A-I-containing lipoproteins  
A:Reference number: A54223; MUID:94162201; PMID:8117655  
A:Accession: C54223  
A:Molecule type: protein  
A:Residues: 'X', 22, 'X', 24, 'X', 26-31, 'X', 33-34 <KUN>  
R:Tenkanen, H.; Lukka, M.; Jaatinen, M.; Metsu, J.; Baumann, M.; Peltonen, L.; Ehnholm  
Arterioscler. Thromb. 11, 851-856, 1991  
A:Title: The mutation causing the common apolipoprotein A-IV polymorphism is a glutamine  
A:Reference number: A61203; MUID:91291788; PMID:2065039  
A:Accession: A61203  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 378-379, 'H', 381-382 <TEN>

C:Comment: ApoA-IV is a major lipoprotein of lymph chylomicrons. In human plasma, it  
C:Comment: ApoA-IV is synthesized primarily in the intestine.  
C:GeneID: 119000  
A:Gene: GDB:AP0A4  
A:Cross-references: GDB:119000; OMIM:107690  
A:Map position: 11q23-11q23  
A:Introns: 17/1, 58/2  
C:Superfamily: apolipoprotein A-I  
C:Keywords: chylomicron; HDL; intestine; lipid binding; lipid transport; lipoprotein  
F1-20/Domains: signal sequence \*status experimental <Sig>  
F:21-396/Product: apolipoprotein A-IV \*status experimental <Mat>

Query Match	17.5%	Score 328	DB 1	Length 396
Best Local Similarity	28.4%	Pred. NO. 8.2e-13		
Matches	96	Conservative	72	Mismatches 152; Indels 18; Gaps 6

QY 5 AAVLTWALALISA----FSATQARKGFWDYFSQTSGD-KGRVEQIHQKMKARE-PATLKD 58

Db 5 AVLTTLVAACGARAEEVSADOVATVMWDYFSQI.SNNAKAEVHLLQKSEITOOINAI.FOD 64

59 SLEODI NNMNKEI EKI RPI SCSEADRI PODBYCMPROIOEEL EEWKARI ODYMAFAUEL V 11

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

03 ABGEVNIYAGDLOKLLVPFAVEHIEKRLANDSEKLEELIGKEDEELKARLLPHANEVSKI 12

```
QY      119 GWNLEGLRQQJ,KPYTMDLMEQVALRVQELQEQI,RVNGEDTKAQLLCGYDEAMALLQ----- 177
```

Db 125 GDNIRELQQRLEPYADQLRTQVNTQAEQLRRQLDPLAQRMERYLTRENADSLQASLRPHAD 18

QY 175 GLOSRVHHHTGKRELFHPYAESLVSGIGRHVQELHRSVAPHAPASPAPLSPCYVLSRK 233

Db 185 ELKAKIDNVEELKGRITPYADEFKVKIDQTVHELRRSLADYAQDTQEKLNHQLGEGITVQ 24

235 LTLKAKALHARIQONIDQIRELSRAFGT-----GTEEGAGDPQMLSEEVQRLOAFR 280

```

Db      245  MKKNAEELKARISAAEELRRLAPLAEDVRGNLKNTGLOKSLAELGGHIDDOVEEER 30
          : | : | | | : : | | : | | : | | : | | : | | : | |

```

0x 290 0D77101AAETRAID0E7FEV0001AP---DDPGHSAF 324

305 **PRIVREKUPNOSTI UOČENI NA POSREDOVANJE ZA 2005**

303 KRYEIGENT NALVQREQLRQLSPHAGDVEGHLE 342

### RESULT 3

apoliipoprotein A-IV precursor - rat

C:Species: *Rattus norvegicus* (Norway rat)  
C:Date: 27-Nov-1985 #sequence revision 27-Nov-1985 #rev change 23-Feb-1986

C:\Access1: A03095; A25214; C24700

R.; Boguski, M.S.; Fishnourdagy, N.; Taylor, J.M.; Gordon, J.L.  
Proc. Natl. Acad. Sci. U.S.A. 81, 5021-5025, 1984

A;Title: Rat apolipoprotein A-IV contains 13 tandem repetitions of a 22-a  
A;Reference number: A03095; MUID:84298074; PMID:6591177

A: Accession: A03095  
A: Molecule type: mRNA

A;Residues: 1-391 <BOG>  
A:Cross-References: CB.W00003, CB.W03431, NID-200040, PIDN.M000000 1, C

R. Boguski, M.S.; Birkenmeier, E.H.; Elshourbagy, N.A.; Taylor, J.M.; Gordon, J.M. *Genetics* 2005, 160, 1133–1142.

J. Biol. Chem. 261, 6398-6407, 1986

A:Reference number: A25214; MUID:86196059; PMID:3009456  
A:Accession: A25214

A: Molecule type: protein  
A: Residues: 1-252 / 0 / 254-391 <BO3>

R;Haddad, I.A.; Ordovas, J.M.; Fitzpatrick, T.; Karathanasis, S.K.

A/Title: Linkage, evolution, and expression of the rat apolipoprotein A-I

A;Reference number: A92558; MUID:87008540; PMID:3020028  
A;Accession: C24700

A: Molecule type: DNA  
A: Residues: 1-252, 'O', 254-391 <HAD>

A: Cross-references: GB:J02588; NID:g202937; PIDN:AAA0747.1; PID:g202941  
Comment: This monograph is a major component of our and that of the

C/Comment: Nine of the thirteen 22-amino acid tandem repeats (each 22-mer

cal, and many of these helices are amphipathic. They may therefore serve as

Downloaded from <http://ajph.org/> on November 10, 2014





## RESULT 9

apolipoprotein A-I precursor [validated] - human

N:Alternate names: APOA-I-2; APOA-I-4; preproapoA-I; proslacylin stabilizing factor

C:Species: Homo sapiens (man)

C:Date: 14-Nov-1983 #sequence, revision 14-Nov-1983 #text, change 08-Dec-2000

C:Accession: A90947; B90947; S02373; A93465; A21147; A93519; B93519; A93472; A94010; A21147

R:Selkammer, J.J.; Procter, A.A.; Frossard, P.; Levy-Wilson, B.

DNA 3, 309-317, 1984

A:Title: Isolation and DNA sequence of full-length cDNA and of the entire gene for human

A:Reference number: A90947; MUID:8502665; PMID:6207999

A:Accession: A90947

A:Molecule type: DNA

A:Residues: 1-267 <SEI>

A:Cross-references: GB:J01038; NID:928769; PIDN:CAA25519.1; PID:9296635

A:Accession: B90947

A:Molecule type: mRNA

A:Residues: 1-267 <SE2>

A:Cross-references: GB:X01038; NID:928769; PIDN:CAA25519.1; PID:9296635

R:Markides, S.C.; Ruiz-Opazo, N.; Hayden, M.; Nusbaum, A.L.; Breslow, J.L.; Zannis, V.I.

Eur. J. Biochem. 173, 465-471, 1988

A:Title: Sequence and expression of Tangier apoA-I gene.

A:Reference number: S02373; MUID:88196137; PMID:3129297

A:Accession: S02373

A:Molecule type: DNA

A:Residues: 1-267 <MAX>

A:Cross-references: EMBL:X07496; NID:928774; PIDN:CAA30377.1; PID:9296729

R:Shoulders, C.C.; Kornblith, A.R.; Munro, B.S.; Baralle, F.E.

Nucleic Acids Res. 11, 2827-2837, 1983

A:Title: Gene structure of human apolipoprotein AI.

A:Reference number: A93465; MUID:83220822; PMID:6406984

A:Accession: A93465

A:Molecule type: DNA

A:Residues: 1-267 <SHO>

A:Cross-references: GB:J00098; GB:J00100; GB:J00101; GB:J03222; GB:K01518; GB

R:Karathanasis, S.K.; Zannis, V.I.; Breslow, J.L.

Proc. Natl. Acad. Sci. U.S.A. 80, 6147-6151, 1983

A:Title: Isolation and characterization of the human apolipoprotein A-I gene.

A:Reference number: A21147; MUID:84016011; PMID:6413973

A:Accession: A21147

A:Molecule type: DNA

A:Residues: 1-267 <KAR>

A:Cross-references: GB:J00098; GB:J03222; NID:917875; PIDN:AA959514.1; PID:9178768

R:Shapiro, C.R.; Sidoli, A.; Shelley, C.S.; Lucero, M.A.; Shoulters, C.C.; Baralle, F.E.

Nucleic Acids Res. 12, 3917-3932, 1984

A:Title: Human apolipoproteins AI, AII, CII and CIII. cDNA sequences and mRNA abundance.

A:Reference number: A93519; MUID:84221405; PMID:6328445

A:Accession: A93519

A:Molecule type: mRNA

A:Residues: 1-267 <SHA>

A:Cross-references: GB:X00566; NID:928765; PIDN:CAA2532.1; PID:9372753

A:Accession: B93519

A:Molecule type: DNA

A:Residues: 1-24 <SH2>

R:Cheung, P.; Chan, L.

Nucleic Acids Res. 11, 3703-3715, 1983

A:Title: Nucleotide sequence of cloned cDNA of human apolipoprotein A-I.

A:Reference number: A93472; MUID:83220772; PMID:6304641

A:Accession: A93472

A:Molecule type: mRNA

A:Residues: 1-267 <CHE>

A:Cross-references: GB:J00098; GB:J00099; GB:J00100; GB:J00101; GB:J03222; GB:K01518; GB

R:Law, S.W.; Brewer Jr., H.B.

Proc. Natl. Acad. Sci. U.S.A. 81, 66-70, 1984

A:Title: Nucleotide sequence and the encoded amino acids of human apolipoprotein A-I mRNA

A:Reference number: A94010; MUID:84119464; PMID:6198645

A:Accession: A94010

A:Molecule type: mRNA

A:Residues: 1-267 <LAW>

A:Cross-references: GB:J00098; GB:J00099; GB:J00100; GB:J00101; GB:J03222; GB:K01518; GB

R:Zannis, V.I.; Karathanasis, S.K.; Neumann, H.T.; Goldberger, G.; Breslow, J.L.

Proc. Natl. Acad. Sci. U.S.A. 80, 2574-2578, 1983

A:Title: Intracellular and extracellular processing of human apolipoprotein A-I: secret

A:Reference number: A21118; MUID:83195100; PMID:6405383

A:Accession: A21118

A:Molecule type: mRNA

A:Residues: 1-24 <ZAN>

R:Brewer Jr., H.B.; Fairwell, T.; Kay, L.; Meng, M.; Roman, R.; Law, S.; Light, J.A.

Biochem. Biophys. Res. Commun. 113, 626-632, 1983

A:Title: Human plasma proapoA-I: isolation and amino-terminal sequence.

A:Reference number: A90112; MUID:83256553; PMID:6409108

A:Accession: A90112

A:Molecule type: protein

A:Residues: 19-27 <BRE>

R:Brewer Jr., H.B.; Fairwell, T.; Lank, A.; Roman, R.; Houser, A.; Bronzert, T.J.

Biochem. Biophys. Res. Commun. 80, 623-630, 1978

A:Title: The amino acid sequence of human APOA-I, an apolipoprotein isolated from h

A:Reference number: A90209; MUID:78123731; PMID:204308

A:Accession: A90209

A:Molecule type: protein

A:Residues: 25-57, 'Q', '59-169, 'OO', '172-267 <HR2>

R:Yui, Y.; Aoyama, T.; Morishita, H.; Takahashi, M.; Takatsu, Y.; Kawai, C.

J. Clin. Invest. 82, 803-807, 1988

A:Title: Serum proslacylin stabilizing factor is identical to apolipoprotein A-I (AI

A:Reference number: A30516; MUID:8831387; PMID:3047170

A:Accession: A30516

A:Molecule type: protein

A:Residues: 25-56 <YUI>

R:Nichols, W.C.; Dwulet, F.E.; Liepnies, J.; Benson, M.D.

Biochem. Biophys. Res. Commun. 156, 762-768, 1988

A:Title: Variant apolipoprotein AI as a major constituent of a human hereditary amy

A:Reference number: A31582; MUID:89050104; PMID:3142462

A:Accession: A31582

A:Molecule type: protein

A:Residues: 25-49, 'R', '51-85, 'D', '87-107 <NIC>

A:Note: Variant sequence from patient with familial amyloidotic polynuropathy type J

R:Manjunath, P.; Marcel, Y.L.; Uma, J.; Seidah, N.G.; Chretien, P.; Chapdelaine, A.

J. Biol. Chem. 264, 16853-16857, 1989

A:Title: Apolipoprotein A-I binds to a family of bovine seminal plasma proteins.

A:Reference number: A34409; MUID:89380318; PMID:2506184

A:Accession: A34409

A:Molecule type: protein

A:Residues: 25-48 <MAN>

R:Stoffel, W.; Binczek, E.

Biol. Chem. Hoppe-Seyler 369, 1055-1063, 1988

A:Title: Structural requirements of human preproapolipoprotein AI for translocation t

A:Reference number: S02737; MUID:89149957; PMID:3228490

A:Accession: S02737

A:Molecule type: protein

A:Residues: 1-6, 'AV', '9, '14V', '12-29 <STO>

A:Note: part of this sequence, including the amino end of the mature protein, was cor

R:Stoffel, W.; Binczek, E.

Biol. Chem. Hoppe-Seyler 372, 481-488, 1991

A:Title: Transient expression of wild type and mutant human apolipoprotein AI in COS

A:Reference number: S16197; MUID:92029676; PMID:1930731

A:Contents: annotation; extension; studies in reference S02737

R:Stoffel, W.; Krueger, E.; Deutzmann, R.

Hoppe-Seyler's Z. Physiol. Chem. 364, 227-237, 1993

A:Title: Cell-free translation of human liver apolipoprotein AI and AII mRNA process

A:Reference number: A19913; MUID:83236195; PMID:6407957

A:Accession: B19913

A:Molecule type: protein

A:Residues: 1-6, 'X', '8-13, 'XXX', '17-18, 'XX', '21, 'X', '23-25, 'X', '27-29 <ST2>

R:Enholm, C.; Bozas, S.E.; Tenkansen, H.; Kirsbaum, L.; Metso, J.; Murphy, B.; Walke

Biochim. Biophys. Acta 1086, 255-260, 1991

A:Title: The apolipoprotein A-I binding protein of placenta and the SP-40,40 protein

A:Reference number: A56815; MUID:92075698; PMID:1742316

A:Accession: A56815

A:Molecule type: protein

A:Residues: 25-31, 'P', '33 <EHN>

A:Experimental source: serum

A:Note: sequence extracted from NCBI backbone (NCBIRP:69759)

A:Note: 32-Tip was also found

R:Kunitake, S.T.; Carilli, C.T.; Lau, K.; Protter, A.A.; Naya-Vigane, J.; Kane, J.P. Biochemistry 33, 1988-1993, 1994  
 A:Title: Identification of proteins associated with apolipoprotein A-I-containing lipoproteins  
 A:Reference number: A54223; MUID:94162201; PMID:8117655  
 A:Accession: A54223  
 A:Molecule type: protein  
 A:Residues: 25-39 <KUN>  
 R:Kobayashi, N.; Koo, C.; Loriau, R.; Guillaume, J.P.; Jacobs, P.; Cravador, A.; He DNA 8, 429-436, 1989  
 A:Title: Production of human recombinant proapolipoprotein A-I in *Escherichia coli*: purification and characterization  
 A:Reference number: 139476; MUID:89377481; PMID:2673706  
 A:Accession: 139476  
 A:Molecule type: mRNA  
 A:Residues: 19-267 <RES>  
 A:Cross-references: GB:M29068; NID:q178774; PIDN:AAA51746.1; PID:q178775  
 R:Higuchi, K.; Law, S.W.; Hoeg, J.M.; Schumacher, U.K.; Meglin, N.; Brewer, H.H. J. Biol. Chem. 263, 18530-18536, 1988  
 A:Title: Tissue-specific expression of apolipoprotein A-I (ApoA-I) is regulated by the 5' nontranscribed region  
 A:Reference number: 139475; MUID:89054040; PMID:3142880  
 A:Accession: 139475  
 A:Molecule type: DNA  
 A:Residues: 1-14 <RE2>  
 A:Cross-references: GB:J04066; NID:q178763; PIDN:AAA51746.1; PID:q553183  
 R:Breslow, J.L. Annu. Rev. Biochem. 54, 699-727, 1985  
 A:Title: Human apolipoprotein molecular biology and genetic variation.  
 A:Reference number: A90042; MUID:85278004; PMID:3896129  
 A:Accession: A90042  
 A:Contents: annotation: review of sequences, variants and gene location  
 R:Hoeg, J.M.; Meng, M.S.; Roman, R.; Fairwell, T.; Brewer Jr., H.B. J. Biol. Chem. 261, 3911-3914, 1986  
 A:Title: Human apolipoprotein A-I. Post-translational modification by fatty acid acylation  
 A:Reference number: A92577; MUID:86140194; PMID:3005308  
 A:Accession: A92577  
 A:Contents: annotation: acylation with palmitate  
 A:Note: an underlined serine or threonine is acylated by fatty acid; the acylating fatty acid is indicated by the letter 'X'  
 R:Law, S.W.; Brewer, H.B. J. Biol. Chem. 260, 12810-12814, 1985  
 A:Title: Tangier disease: The complete mRNA sequence encoding for preproapo A-I.  
 A:Reference number: 155236; MUID:86008382; PMID:2995392  
 A:Accession: 155236  
 A:Molecule type: mRNA  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Residues: 1-143, 'D', 145-267 <RE3>  
 A:Cross-references: GB:M11791; NID:q178776; PIDN:AAA55545.1; PID:q178777  
 C:Comment: Apolipoprotein A-I precursor is synthesized in the liver and small intestine.  
 C:Genetics: HDL in plasma.  
 A:Gene: GDB:AP0A1  
 A:Cross-references: GDB:119684; OMIM:107680  
 A:Map position: 11q23.3-11q23.3  
 A:Introns: 15/1: 67/2  
 C:Function: participates in the reverse transport of cholesterol from tissues to the liver  
 C:Superfamily: apolipoprotein A-I  
 C:Keywords: atherosclerosis; cholesterol metabolism; HDL; intestine; lipid binding; lipid transport  
 F:1-18/Domain: signal sequence #status experimental <SIG>  
 F:19-24/Domain: propeptide #status experimental <PRO>  
 F:25-267/Product: apolipoprotein A-I #status experimental <MAT>  
 Query Match 10.7% Score 200; DB 1; Length 267;  
 Best Local Similarity 22.2% Pred. No. 2.9e-05;  
 Matches 74; Conservative 61; Mismatches 106; Indels 92; Gaps 10;

DB 119 AKVQYPLDDFOKKQEMELRYQKVEPLRAELQEGAROKLHLEOKLSPLGEEMR----- 173  
 QY 165 GVDEAMALLQGLQSRVYHHTGRFKELFHPYASLVSGICRHVOELHRSVAPAPASPARL 224  
 DB 174 --DRARA-----HVDALRTHLAPYSDELRLQRL 198  
 QY 225 SRCVOYLSKRLTKAKALHARTQONLDRELRSRFACTGTEEGAGPPQMLSEVRR 284  
 DB 199 AARLEALKNGCARLAEXHAKATEHLSTLSEKAKAL-----EDLRQ 241  
 QY 285 ----LQAFRODTYLOIAFTRAIDQTEFVQOO 313  
 DB 242 LLPVLESFK-----VSFLSLEETTKLNTQ 267  
 KRSUN 10 JS0079  
 apolipoprotein A-I precursor - baboon  
 C:Species: Papio sp. (baboon)  
 C:Date: 31-Mar-1992 #sequence-revision 31-Mar-1992 #text-change 13-Jun-1997  
 C:Accession: JS0079  
 R:Hixson, J.H.; Bornstein, S.; Cox, L.A.; Rainwater, D.L.; Vandeberg, J.L. Gene 74, 483-490, 1988  
 A:Title: The baboon gene for apolipoprotein A-I: characterization of a cDNA clone and protein sequence  
 A:Reference number: JS0079; MUID:89232739; PMID:2907746  
 A:Accession: JS0079  
 A:Molecule type: mRNA  
 A:Residues: 1-267 <HX>  
 A:Experimental source: liver  
 C:Comment: This protein is the principal protein component of high density lipoprotein  
 C:Comment: This protein serves as a cofactor for lecithin cholesterol acyltransferase  
 C:Comment: This protein contains a region of repeated amino acids which form amphipathic alpha-helices  
 A:Gene: apoA1  
 C:Superfamily: apolipoprotein A-I  
 C:Keywords: HDL; lipid binding; lipoprotein  
 F:1-18/Domain: signal sequence #status predicted <SIG>  
 F:19-267/Product: apolipoprotein A-I #status predicted <LAI>  
 F:123-144, 145-166, 167-188, 189-210, 211-232, 233-254/Region: tandem repeats  
 Query Match 10.3% Score 193; DB 2; Length 267;  
 Best Local Similarity 22.2% Pred. No. 7.6e-05;  
 Matches 74; Conservative 61; Mismatches 106; Indels 92; Gaps 10;



C>Date: 30-Sep-1989 #sequence revision 19-Oct-1995 #text\_change 22-Jun-1999

C:Accession: A26529; A26627; S23135; A57766

R:Polites, H.G.; Melchior, G.W.; Castile, C.K.; Marottili, K.R.

Gene 49, 103-110, 1986

A:Title: The primary structure of cynomolgus monkey apolipoprotein A-I deduced from the

A:Reference number: A26529; PMID:87191989; PMID:3106152

A:Accession: A26529

A:Molecule type: mRNA

A:Residues: 1-267 <POL>

A:Cross-references: GB:M15411; NID:g342074; PIDN:AAA36834.1; PID:g342075

R:Heberly, P.N.; Bausserman, L.L.; Lynch, K.M.; Saritelli, A.L.; Kantor, M.A.; Nicolosi,

Biochemistry 26, 1457-1463, 1987

A:Title: Homologues of the human C and A apolipoproteins in the Macaca fascicularis (cyr

A:Reference number: A26627; PMID:87185451; PMID:3105581

A:Accession: A26627

A:Molecule type: protein

A:Residues: 25-48 <HER>

R:Murray, R.W.; Marottili, K.R.

Biochim. Biophys. Acta 1131, 207-210, 1992

A:Title: Nucleotide sequence of the cynomolgus monkey apolipoprotein A-I gene and corres

A:Reference number: S23135; PMID:92305062; PMID:1610902

A:Accession: S23135

A:Molecule type: DNA

A:Residues: 1-12, 'L', 14-267 <MUR>

A:Cross-references: GB:M83242; NID:g342070; PIDN:AAA36832.1; PID:g342071

R: Sorci, Thomas, M.; Kearns, M.W.

J. Biol. Chem. 266, 18045-18050, 1991

A:Title: Transcriptional regulation of the apolipoprotein A-I gene.

A:Reference number: A57766; PMID:92011532; PMID:1917942

A:Accession: A57766

A:Molecule type: DNA

A:Residues: 1-110 <RES>

A:Cross-references: GB:M69223; NID:g342066; PIDN:AAA36831.1; PID:g553820

C:Comment: The precursor is synthesized in the liver and small intestine. The propeptide

C:Comment: ApoA-I makes up about 70% of the protein of the high density lipoproteins (HDL)

C:Comment: promoting cholesterol efflux from tissues and by acting as a cofactor for the lecithin

C:Genetics:

A:Insertions: 15/1; 67/2

C:Superfamily: apolipoprotein A-I

C:Keywords: cholesterol metabolism; HDL; intestine; lipid transport; plasma; tandem repe

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-24/Domain: propeptide #status predicted <PRO>

F:25-267/Product: apolipoprotein A-I #status predicted <MAT>

Query Match

Best Local Similarity 10.0%; Score 187; DB 1; Length 267;

Matches 75; Conservative 61; Mismatches 105; Indels 92; Gaps 11;

5 AAVITMALALSAFSAQARQKGFV--DYFSQSGD-----KGRVQDIN 45

3 ATVTTLAVLTL--TGSQARH-FWQODEPQTPMDRVKDLVTYVVALKDSKDYVSQFE 58

46 QOKKARE-PATLKDSLEODLNMMNKFLKRLPLSGSEAPRLPDVPGMRLOLEEEVK 104

59 GSAIGKQLNLKLDNMNSVSTSVKLRQGLPVTGFEMDMNLEKTEGRLQGMDSMDLEVK 118

105 ARLOPYMAEHELVGMNLEGLRQOLKPYTMDLMEQVALRVOELRQVGEDTKAQLLG 164

119 AKVOPYLDDDFQKMQEEMELTRQKVEPLRAELHGTGRKLEHELEKSPLOEEVR---- 173

165 GVDWAMLLQGLQSRVYVHTGRFKEFLHPYAESLVSGIRHVOELHRSVAPHAAPARL 224

174 --DVARA-----HVDALRTHLAFYSDELKQRL 198

225 SRCVQVLSRKLTLLAKALHARIQNDLQRELSRAFGTGEAGCPDQMLSEVQR 284

199 AARLEAKKENGARLAELHAKASEHLSTLSEKAKPAL-----EDLRQG 241

285 ----LQAFRODTYQLIAFTPAIDQETEEYQGO 313

242 LLPVLESFK-----VSFLSALEETVTKLSTQ 267

RESULT 12

apolipoprotein A-I precursor (clone 22ap A1) - rabbit

C:Species: Oryctolagus cuniculus (domestic rabbit)

C>Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 22-Jun-1999

C:Accession: S06064

R:Paraskeopoulou, T.B.; Kritis, A.; Zannis, V.

submitted to the EMBL Data Library, July 1989

A:Reference number: S06064

A:Accession: S06064

A:Molecule type: mRNA

A:Residues: 1-266 <PAR>

A:Cross-references: EMBL:X15908; NID:g1457; PIDN:CAA34024.1; PID:g1458

C:Comment: This protein is synthesized in the small intestine.

C:Superfamily: apolipoprotein A-I

C:Keywords: cholesterol metabolism; HDL; intestine; lipid transport; plasma; tandem r

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-24/Domain: propeptide #status predicted <PRO>

F:25-266/Product: apolipoprotein A-I #status experimental <MAT>

Query Match

Best Local Similarity 9.9%; Score 185; DB 1; Length 266;

Matches 61; Conservative 61; Mismatches 122; Indels 28; Gaps 6;

5 AAVITMALALSAFSAQAR-----KGFMD-----YFSQSGDKGR--VEQ1HQQ 47

3 AVTTLAVLTL--TGSQARHFWQODEPQTPMDRVKDLVTYVVALKDSKDYVSQFE 59

48 KMARE-PATLKDSLEODLNMMNKFLKRLPLSGSEAPRLPDVPGMRLOLEEEVKAR 106

60 AFGKOLMKLDNMNSVSTSVKLRQGLPVTGFEMDMNLEKTEGRLQGMDSMDLEVK 119

107 LQPYMAEHELVGMNLEGLRQOLKPYTMDLMEQVALRVOELRQVGEDTKAQLLG 166

120 VQPLDFQKMQEEMELTRQKVEPLRAELHGTGRKLEHELEKSPLOEEVR----- 172

167 DEAMWLLQGLQSRVYVHTGRFKEFLHPYAESLVSGIRHVOELHRSVAPHAAPARL 226

173 DSARTHTDTRTKAPRNSNELOGLARLESIRGEGASLAEOAKAREHLVSLSEKAR 232

227 CVQVLSRKLTLLAKALHARIQNDLQRELS 258

233 ALMDLRQGLLPLVLESFKASVQNVLDFAVTKLIN 264

apolipoprotein A-I precursor - chicken

C:Species: Gallus gallus (chicken)

C>Date: 31-Mar-1990 #sequence\_revision 25-Aug-1995 #text\_change 22-Jun-1999

C:Accession: JH0471; A29657; S01453; S28888; A29616; S00187; S10973; S11023; I50156

R:Bhattacharya, N.; Chattopadhyay, R.; Hirsch, A.; Banerjee, D.

Gene 104, 163-168, 1991

A:Title: Isolation, characterization and sequencing of the chicken apolipoprotein-AI

A:Reference number: JH0471; PMID:92009209; PMID:1916289

A:Accession: JH0471

A:Molecule type: DNA

A:Residues: 1-264 <BHA>

R:Byrnes, L.; Luo, C.C.; Li, W.H.; Yang, C.; Chan, L.

Biochim. Biophys. Res. Commun. 148, 485-492, 1987

A:Title: Chicken apolipoprotein A-I: cDNA sequence, tissue expression and evolution.

A:Reference number: A29657; PMID:88049703; PMID:3118875

A:Accession: A29657

A:Molecule type: mRNA

A:Residues: 1-264 <BYR>

A:Cross-references: EMBL:M17961; NID:g211147; PIDN:AAA48593.1; PID:g211148

R:Rajavashisth, T.B.; Dawson, P.A.; Williams, D.L.; Shackelford, J.E.; Leberitz, H.; I

J. Biol. Chem. 262, 7058-7065, 1987

A:Title: Structure, evolution, and regulation of chicken apolipoprotein A-I.

A:Reference number: S01453; PMID:87223301; PMID:3108248

A:Accession: S01453

A:Molecule type: mRNA



A:Residues: 1-15, '1', '17-147, 'K', '149-264 <RAJ>  
A:Cross-references: GB:M25359; EMBL:J02739; NID:g211145; P1DN:AAA48592.1; P1D:g211146  
A:Accession: S28888  
A:Molecule type: protein  
A:Residues: 25-44; 230-256 <RA2>  
R:Ferrari, S.; Tarugi, P.; Drusiani, E.; Calandra, S.; Fregni, M.  
Gene 60, 39-46, 1987  
A:Title: The complete sequence of chick apolipoprotein A-I mRNA and its expression in the  
A:Reference number: A29616; MUID:88152500; PMID:3126099  
A:Accession: A29616  
A:Molecule type: mRNA  
A:Residues: 1-264 <FRK>  
A:Cross-references: EMBL:M18746; NID:g211149; P1DN:AAA48594.1; P1D:g211150  
R:Yang, C.Y.; Gu, Z.W.; Patsch, W.; Weng, S.A.; Kim, T.W.; Chan, L.  
FEBS Lett. 224, 261-266, 1987  
A:Title: The complete amino acid sequence of proapolipoprotein A-I of chicken high densi  
A:Reference number: S00187; MUID:88083548; PMID:3121386  
A:Accession: S00187  
A:Molecule type: protein  
A:Residues: 19-264 <YAN>  
R:Banerjee, D.; Mukherjee, T.K.; Redman, C.M.  
J. Cell Biol. 101, 1219-1226, 1985  
A:Title: Biosynthesis of high density lipoprotein by chicken liver: intracellular transp  
A:Reference number: S10973; MUID:86008443; PMID:3930506  
A:Accession: S10973  
A:Molecule type: protein  
A:Residues: 19, 'H', '21-23, 'Q', '25-44 <BAN>  
A:Note: the sequence from Fig. 4 is inconsistent with that from Table II in having 41-Th  
R:Shackelford, J.E.; Leberer, H.G.  
J. Biol. Chem. 258, 7175-7180, 1983  
A:Title: Synthesis and secretion of apolipoprotein A-I by chick breast muscle.  
A:Reference number: S11023; MUID:83213468; PMID:6406496  
A:Accession: S11023  
A:Molecule type: protein  
A:Residues: 25-44 <SHM>  
R:Lamson-Pava, S.; Ferrari, R.; Rajavashisth, T.B.; Lusis, A.J.; Karathanasis  
J. Lipid Res. 33, 831-842, 1992  
A:Title: Evolutionary distinct mechanisms regulate apolipoprotein A-I gene expression: D  
A:Reference number: I50156; MUID:92381402; PMID:1512510  
A:Accession: I50156  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-15, '1', '17-264 <IAM>  
A:Cross-references: GB:M66012; NID:g211158; P1DN:AAA48597.1; P1D:g211159  
C:Comment: This protein is synthesized only in the liver in mammals, whereas in chicken  
C:Comment: This protein is a major component of the high density lipoproteins in plasma.  
C:Genetics:  
A:Gene: ApoA1  
A:Introns: 15/1; 66/2  
C:Superfamily: apolipoprotein A-I  
C:Keywords: cholesterol metabolism; HDL; intestine; lipid binding; lipid transport; lipo  
F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-24/Domain: propeptide #status experimental <PRO>  
F:25-264/Product: apolipoprotein A-I #status experimental <MAT>  
Query Match 9.7%; Score 183; DB 1; Length 264;  
Best local similarity 21.1%; Pred. No. 0.0003;  
Matches 70; Conservative 62; Mismatches 113; Indels 86; Gaps 9;

DB 178 -----HVEELRKNLAPYSDELROK. 197  
QY 225 SRGVLSRKLTLLKAKALHRIQNLDOELRELSAFAGTGTGEEAGDPOMLSEVQR 284  
DB 198 SOKLEIREKGIPOASEYQAKVMEQSLNREKMT -----PLVGEFRER 240  
QY 285 LQAFRODTYLOIAAFTRAIDQETFEVQOIA 315  
DB 241 LTPYENLKNRLISF-----LDELQNSVA 264  
RESULT 14  
LRR31B  
apolipoprotein A-I precursor (clone pRHA-502) - rabbit  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text-change 22-Jun-1999  
C:Accession: S00230; S20557  
R:Pan, T.C.; Hao, Q.L.; Yamin, T.T.; Dai, P.H.; Chen, H.S.; Chen, S.L.; Kroon, P.A.;  
Kur, J. Biochem. 170, 99-104, 1987  
A:Title: Rabbit apolipoprotein A-I mRNA and gene: evidence that rabbit apolipoprotein  
A:Reference number: S00230; MUID:88082866; PMID:3121329  
A:Accession: S00230  
A:Molecule type: mRNA  
A:Residues: 1-265 <PAN>  
A:Cross-references: EMBL:X06658; NID:g1461; P1DN:CAA29857.1; P1D:g1462  
A:Note: the authors translated the codon AGC for residue 174 as Arg  
A:Accession: S20557  
A:Molecule type: DNA  
A:Residues: 1-17, 'R', '19-44, 'I', '46-122, 'Y', '124-146, 'V', '148-265 <PAN2>  
A:Cross-references: EMBL:X06659; NID:g1459; P1DN:CAA29858.1; P1D:g1460  
C:Comment: This protein is synthesized in the small intestine.  
C:Comment: This protein is a major component of the high density lipoproteins in plas  
C:Genetics:  
A:Introns: 15/1; 66/2  
C:Superfamily: apolipoprotein A-I  
C:Keywords: cholesterol metabolism; HDL; intestine; lipid transport; plasma; tandem r  
F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-24/Domain: propeptide #status experimental <PRO>  
F:25-265/Product: apolipoprotein A-I #status experimental <MAT>  
Query Match 9.6%; Score 180; DB 1; Length 265;  
Best local similarity 21.6%; Pred. No. 0.00046;  
Matches 71; Conservative 57; Mismatches 114; Indels 86; Gaps 8;

C>Date: 05-Jan-1996 #sequence-revision 23-Aug-1997 #text-change 13-Aug-1999  
 A:Accession: I45853; A56858; A34649  
 R:O'Huigin, C.; Chan, L.; Li, M.  
 M: Biol. Evol. 7, 327-339, 1990  
 A:Title: Cloning and sequencing of bovine apolipoprotein A-I cDNA and molecular evolution  
 A:Reference number: I45853; M:ID:90348478; PMID:2117227  
 A:Accession: I45853  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-265 <OXH>  
 A:Cross-references: GB:M35870; M:ID:9162677; P:ID:AAA30381.1; P:ID:9162678  
 R:Sparrow, D.A.; Lee, B.R.; Laplaud, P.M.; Aubolton, S.; Baughart, D.; Chapman, M.J.; Gd  
 Biochim. Biophys. Acta 1123, 145-150, 1992  
 A:Title: Plasma lipid transport in the pre-ruminant calf, Bos spp: primary structure of  
 A:Reference number: A56858; M:ID:92153895; PMID:1739745  
 A:Accession: A56858  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 19-184, 'QL', 187-265 <SPA>  
 A:Experimental source: Friesian-Holstein male calves aged 2-4 weeks  
 A:Note: sequence extracted from NCHI backbone (NCBI:83520)  
 R:Aubolton, S.; Sparrow, D.A.; Beaubatie, L.; Baughart, D.; Sparrow, J.T.; Laplaud, P.M.  
 Biochem. Biophys. Res. Commun. 166, 833-839, 1990  
 A:Title: Characterization and amino-terminal sequence of apolipoprotein AI from plasma  
 A:Reference number: A34649; M:ID:90147795; PMID:2105728  
 A:Accession: A34649  
 A:Molecule type: protein  
 A:Residues: 25-70 <AUB>  
 A:Experimental source: Friesian-Holstein male calves aged 2-4 weeks  
 C:Superfamily: Apolipoprotein A-I  
 C:Keywords: lipid binding; lipoprotein  
 F:1-18/Domain: signal sequence #status predicted <SIG>  
 F:19-265/Product: apolipoprotein A-I #status experimental <MAT>

Query Match 9.5%; Score 178; DB 2; length 265;  
 Best Local Similarity 22.3%; Pred. No. 0.00061;  
 Matches 75; Conservative 57; Mismatches 103; Indels 102; Gaps 10;

```

QY 5 AAVLTALALISAFSAQAR-----KGFMDYFSOTSCDKGR--VEQIHQ 47
DB 3 AAVLTALAVLFL---TGSQARHPWQODDPOSSMDRVKDFATVYVEAIKDSGRDYVAQFEAS 59
QY 48 KMARE-PATLKDSLQDPLNNMKFLKRLPLSGSEAPRLPDQPYGMRRQLQEELEEVKAR 106
DB 60 ALCKQLNLKLLDMWDTLASTLSKYVRDLGPTQGFMDNLEKETASLRQEMHKDLEEVKOK 119
QY 107 LQPYMAEHLVGMNLEGLRQQLKPYTMDLMKQVALKVFLOQLRVYGFRTKAQLIGCV 166
DB 120 VQPYLDEFQKKHVEVEIYRQKVAPLGEEFEREGARQKVOELQDKL-----164
QY 167 DEAMALIQCLOSRYVHHTGRFKEI.FHPYAESIYSCIGRHQOELHRSVAPAPASPARLSR 226
DB 165 -----SPLAQELNDRARAHAYETLRQHVAPYS-----190
QY 227 CVOVLSRKLTLKAKAL-----HARIQNLIDLREELSRFACTGTEGAGDPDPOM 276
DB 191 --DDLQRQLTARLEALKEGGSIAEYHAKASPOLKALGEEK-----AKP-----231
QY 277 LSEEVRORLQAFRODYTLQIAFTRAIDQETEEVQOO 313
DB 232 VLEDLRQGLLPVLESLSKVSILA--AIDEASKKLINAO 265

```

Search completed: January 23, 2003, 17:46:45  
 Job time : 39 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 23, 2003, 17:34:16 (Search time 20 seconds)

(without alignments)  
759.017 Million cell updates/sec

Title: US-09-842-364-3

Perfect score: 1877

Sequence: 1 MASMAAVLTWALALSAFSA.....LWEDITHSLHQGHSLGDP 366

Scoring table: BLOSUM62

Gapop 10.0, Gapept 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Swissprot\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	350	18.6	382	AP04_PIG	O46409 sus scrofa
2	333	17.7	429	AP04_MUCFA	P33621 macaca fasc
3	327	17.4	396	AP04_HUMAN	P06727 homo sapien
4	315	16.8	401	AP04_PAPAN	Q28758 papio anubi
5	311.5	16.6	391	AP04_RAT	P02651 rattus norv
6	298.5	15.9	395	AP04_MOUSE	P06728 mus musculu
7	200	10.7	267	AP01_HUMAN	P02647 homo sapien
8	193	10.3	267	AP01_MACFA	P15668 macaca fasc
9	185	9.9	266	AP01_RABIT	P09809 oryctolagus
10	183	9.7	264	AP01_CHICK	P08250 gallus gall
11	179	9.5	262	AP02_ONCMY	O57524 oncorhynchu
12	176	9.4	265	AP01_BOVIN	P15497 bos taurus
13	175.5	9.4	266	AP01_CANFA	P02648 canis famli
14	171.5	9.1	264	AP01_CANAPL	O42296 anas platyr
15	170	9.1	265	AP01_PIG	P18648 sus scrofa
16	168	9.0	317	AP01_MACFA	P10517 macaca fasc
17	168	8.9	317	AP01_PAPAN	P05770 papio anubi
18	167	8.9	265	AP01_TUPGB	O18759 lumaia gils
19	165	8.8	317	AP01_PIG	P18650 sus scrofa
20	162	8.6	281	AP01_BRARE	O42364 brachydanto
21	161	8.6	264	AP01_MOUSE	O00023 mus musculu
22	160	8.5	262	AP01_SALTR	O91488 salmo trutt
23	156.5	8.3	3674	AP01_HUMAN	O9nc66 homo sapien
24	155	8.3	264	AP01_COTJA	P32918 coturnix co
25	155	8.3	311	AP01_MOUSE	P08826 mus musculu
26	154.5	8.2	312	AP01_RAT	P02650 rattus norv
27	153	8.2	317	AP01_HUMAN	P02649 homo sapien
28	152.5	8.1	298	AP01_CAVPO	P23529 cavia porce
29	152.5	8.1	305	AP01_CANFA	P18649 canis famli
30	152.5	8.1	986	GM13_RAT	O62839 rattus norv
31	150.5	8.0	316	AP01_BOVIN	O03247 bos taurus
32	149	7.9	259	AP01_RAT	P04639 rattus norv
33	149	7.9	311	AP01_RABIT	P18287 oryctolagus

34	148	7.9	262	1	AP01_ONCMY	O57523 oncorhynchu
35	143.5	7.6	1940	1	MYH3_HUMAN	P11055 homo sapien
36	141	7.5	3678	1	DMD_MOUSE	P11531 mus musculu
37	139.5	7.4	258	1	AP01_SALSA	P27007 salmo salar
38	138	7.4	1940	1	MYH3_RAT	P12847 rattus norv
39	137	7.3	262	1	AP01_BRARE	O42363 brachydanto
40	136.5	7.3	848	1	AP01_DIRIM	P13392 diffractaria
41	136.5	7.3	1962	1	MYSA_DROME	P05661 drosophila
42	136.5	7.3	3680	1	DMD_CANFA	O97592 canis famli
43	136.5	7.3	3430	1	ACF7_HUMAN	O9upn3 homo sapien
44	136	7.2	1937	1	MYH8_HUMAN	P13535 homo sapien
45	134.5	7.2	3685	1	DMD_HUMAN	P11532 homo sapien

## ALIGNMENTS

```

RESULT 1
AP04_PIG
ID AP04_PIG STANDARD: PRT; 382 AA.
AC O46409:
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Apolipoprotein A-IV precursor (Apo-AIV).
GN APOA4.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN NCBI_Taxid=9823;
RP SEQUENCE FROM N.A.
RA Osada J., Iturzaide M., Calleja L., Gonzalez N., Pineiro A.;
RL Submitted (DEC-1997) to the EMBL/Genbank/DBD databases.
CC -1- FUNCTION: MAY HAVE A ROLE IN CHYLOMICRONS AND VLDL SECRETION AND
CC CATABOLISM. REQUIRED FOR EFFICIENT ACTIVATION OF LIPOPROTEIN
CC LIPASE BY APOC-II; POTENT ACTIVATOR OF LCAT. APOA-IV IS A MAJOR
CC COMPONENT OF HDL AND CHYLOMICRONS.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: SECRETED IN PLASMA.
CC -1- DOMAIN: NINE OF THE THIRTEEN 22-AMINO ACID TANDEM REPEATS (EACH
CC 22-MR IS ACTUALLY A TANDEM ARRAY OF TWO, A AND B, RELATED) 11-
CC MERS) OCCURRING IN THIS SEQUENCE ARE PREDICTED TO BE HIGHLY ALPHA-
CC HELICAL, AND MANY OF THESE HELICES ARE AMPHIPATHIC. THEY MAY
CC THEREFORE SERVE AS LIPID-BINDING DOMAINS WITH LECITHIN:CHOLESTEROL
CC ACITRANSFERASE (LCAT) ACTIVATING ABILITIES.
CC -1- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL: AJ222966; CA11020.1;
DR HSSP: P32851; 1BR0.
DR InterPro: IPR000074; Apolipoprotein.
KW Plasma; lipid transport; HDL; Chylomicron; Repeat; Signal.
FT SIGNAL 1 20
FT CHAIN 21 382
FT DOMAIN 33 330
FT REPEAT 33 34
FT REPEAT 60 81
FT REPEAT 82 103
FT REPEAT 115 136
FT REPEAT 137 158
FT REPEAT 159 180
FT REPEAT 181 202
FT REPEAT 203 224
FT REPEAT 225

```



CC Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 KC TISSUE=Intestine; PubMed-2930771;  
 RX MEDLINE-89194198;  
 RA Yang C., Gu Z.M., Xiong W., Rosseneu M., Yang H.X., Lee B.M.,  
 RA Goto A.M. Jr., Chan L.,  
 RT "The primary structure of human apolipoprotein A-IV.";  
 RL Biochim. Biophys. Acta 1002:231-237(1989).  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-87041474; PubMed-3095836;  
 RA Karathanasis S.K., Oeltgen P., Haddad I.A., Antonarakis S.E.;  
 RT "Structure, evolution, and polymorphisms of the human apolipoprotein  
 A4 gene (APOA4)."  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:8457-8461(1986).  
 RN [13]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-86296629; PubMed-3755616;  
 RA Karathanasis S.K., Yunis I.;  
 RT "Structure, evolution, and tissue-specific synthesis of human  
 apolipoprotein AIV.";  
 RL Biochemistry 25:3962-3970(1986).  
 RN [14]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-87250378; PubMed-3036793;  
 RA Elshourbagy N.A., Walker D.W., Patk Y.K., Hognuski M.S., Freeman M.,  
 RA Gordon J.I., Taylor J.M.;  
 RT "Structure and expression of the human apolipoprotein A-IV gene.";  
 RL J. Biol. Chem. 262:7973-7981(1987).  
 RN [15]  
 RP SEQUENCE OF 21-396 FROM N.A.  
 RX MEDLINE-86111885; PubMed-3080432;  
 RA Elshourbagy N.A., Walker D.W., Hognuski M.S., Gordon J.I., Taylor J.M.;  
 RT "The nucleotide and derived amino acid sequence of human  
 apolipoprotein A-IV mRNA and the close linkage of its gene to the  
 genes of apolipoproteins A-I and C-III.";  
 RL J. Biol. Chem. 261:1998-2002(1986).  
 RN [16]  
 RP SIGNAL SEQUENCE CLEAVAGE SITE.  
 RX MEDLINE-84161950; PubMed-6706947;  
 RA Gordon J.I., Bisgaler C.L., Sims H.F., Sachdev O.P., Glickman R.M.,  
 RA Strauss A.W.;  
 RT "Biosynthesis of human preapolipoprotein A-IV.";  
 RL J. Biol. Chem. 259:468-474(1984).  
 RN [17]  
 RP REVIEW ON POLYMORPHISM.  
 RA Lohse P., Brewer H.B. Jr.;  
 RT "Genetic polymorphism of apolipoprotein A-IV.";  
 RL Curr. Opin. Lipidol. 2:90-95(1991).  
 RN [18]  
 RP VARIANT A-IV\*2.  
 RX MEDLINE-90277616; PubMed-2351649;  
 RA Lohse P., Kindt M.R., Rader D.J., Brewer H.B. Jr.;  
 RT "Genetic polymorphism of human plasma apolipoprotein A-IV is due to  
 nucleotide substitutions in the apolipoprotein A-IV gene.";  
 RL J. Biol. Chem. 265:10061-10064(1990).  
 RN [19]  
 RP VARIANTS A-IV\*0 AND A-IV\*3.  
 RX MEDLINE-90324273; PubMed-1973689;  
 RA Lohse P., Kindt M.R., Rader D.J., Brewer H.B. Jr.;  
 RT "Human plasma apolipoproteins A-IV-0 and A-IV-3. Molecular basis for  
 two rare variants of apolipoprotein A-IV-1.";  
 RL J. Biol. Chem. 265:12734-12739(1990).  
 RN [10]  
 RP VARIANTS.  
 RX MEDLINE-91310615; PubMed-1677358;  
 RA Lohse P., Kindt M.R., Rader D.J., Brewer H.B. Jr.;  
 RT "Three genetic variants of human plasma apolipoprotein A-IV: apoA-IV-  
 1(Thr-347->Ser), apoA-IV-0(lys-167->Gln, Gln-360->His), and apoA-IV-  
 3(Gln-165->Lys)."  
 RL J. Biol. Chem. 266:13513-13518(1991).

RN [11]  
 RP ERRATUM.  
 RA Lohse P., Kindt M.R., Rader D.J., Brewer H.B. Jr.;  
 RL J. Biol. Chem. 266:19866-19866(1991).  
 RN [12]  
 RP VARIANT MET-13.  
 RX MEDLINE-92238494; PubMed-1349197;  
 RA von Eckardstein A., Funke H., Schulte M., Erten M., Schulte H.,  
 RA Assmann G.;  
 RT "Nonsynonymous polymorphic sites in the apolipoprotein (apo) A-IV  
 gene are associated with changes in the concentration of apo B- and  
 apo A-I-containing lipoproteins in a normal population.";  
 RL Am. J. Hum. Genet. 50:1115-1128(1992).  
 RN [13]  
 RP VARIANT SER-147.  
 RX MEDLINE-92144647; PubMed-1737067;  
 RA Tenkanen H., Koskinen P., Melsö J., Haumann M., Jukka M.,  
 RA Kauppinen-Makelin R., Kontula K., Taskiran M.R., Manttari M.,  
 RA Manninen V., Ehnholm G.;  
 RT "A novel polymorphism of apolipoprotein A-IV is the result of an  
 asparagine to serine substitution at residue 127.";  
 RL Biochim. Biophys. Acta 1138:27-33(1992).  
 RN [14]  
 RP VARIANT A-IV\*5.  
 RX MEDLINE-93138374; PubMed-1487136;  
 RA Kamboh M.I., Williams E.R., Law J.C., Aston C.E., Bunker C.H.,  
 RA Perrelli R.E., Politz W.S.;  
 RT "Molecular basis of a unique African variant (A-IV 5) of human  
 apolipoprotein A-IV and its significance in lipid metabolism.";  
 RL Genet. Epidemiol. 9:379-388(1992).  
 RN [15]  
 RP VARIANT BUDAPEST-2 LYS-44; BUDAPEST-1 CYS-305 AND SER-367.  
 RX MEDLINE-95245341; PubMed-7728150;  
 RA Menzel H.J., Dieplinger H., Sandholzer C., Karadi I., Utermann G.,  
 RA Caszlar A.;  
 RT "Apolipoprotein A-IV polymorphism in the Hungarian population: gene  
 frequencies, effect on lipid levels, and sequence of two new  
 variants.";  
 RL Hum. Mutat. 5:58-65(1995).  
 RN [16]  
 RP VARIANTS FCHE, SHAPLE SER-161; LFU-178 AND GLN-264.  
 RX MEDLINE-97114287; PubMed-8956036;  
 RA Deeb S.S., Nevil D.N., Iwasaki L., Brunzell J.D.;  
 RT "Two novel apolipoprotein A-IV variants in individuals with familial  
 combined hyperlipidemia and diminished levels of lipoprotein lipase  
 activity.";  
 RL Hum. Mutat. 8:319-325(1996).  
 CC -1- FUNCTION: MAY HAVE A ROLE IN CHYLOMICRONS AND VLDL SECRETION AND  
 CATABOLISM. REQUIRED FOR EFFICIENT ACTIVATION OF LIPOPROTEIN  
 LIPASE BY APOC-II: POTENT ACTIVATOR OF LCAT. APOA-IV IS A MAJOR  
 COMPONENT OF HDL AND CHYLOMICRONS.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- TISSUE SPECIFICITY: SYNTHESIZED PRIMARILY IN THE INTESTINE.  
 CC -1- DOMAIN: NINE OF THE THIRTEEN 22-AMINO ACID TANDEM REPEATS (EACH  
 22-MER IS ACTUALLY A TANDEM ARRAY OF TWO, A AND B, RELATED 11-  
 MERS) OCCURRING IN THIS SEQUENCE ARE PREDICTED TO BE HIGHLY ALPHA-  
 HELICAL, AND MANY OF THESE HELICES ARE AMPHIPATHIC. THEY MAY  
 THEREFORE SERVE AS LIPID-BINDING DOMAINS WITH LECITHIN:CHOLESTEROL  
 ACYLTRANSFERASE (LCAT) ACTIVATING ABILITIES.  
 CC -1- POLYMORPHISM: EIGHT ALLELES HAVE BEEN CHARACTERIZED (APOA-IV-0  
 TO APOA-IV\*7). APOA-IV\*1 IS THE MAJOR ALLELE (90%), IV\*2 IS ALSO  
 COMMON (8%), THE OTHERS ARE RARE ALLELES.  
 CC -1- DISEASE: DEFECTS IN APOA4 MAY CONTRIBUTE, ALONG WITH DEFECTS IN  
 OTHER GENES OR ENVIRONMENTAL FACTORS, TO THE DEVELOPMENT OF  
 FAMILIAL COMBINED HYPERLIPIDEMIA (FCHL).  
 CC -1- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial





OY 280 EVRORLQAFRODTYQLIAAFTRALIDETEPVOOQLAPPPGHSAPAFEPQOTDSKV--- 336  
 DB 295 QLDQOYEVERRAVERPLGDFENALVVOOMEKFRQOLG-----SDSGDVESH 339  
 OY 337 LSKLQARLDLWEDITHSLHDG 359  
 DB 340 LSFLEKRLREKVSFMSFTLQKKG 362

RESULT 6  
 APOA4\_MOUSE STANDARD; PRT; 395 AA.  
 AC P06728;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Apolipoprotein A-IV precursor (Apo-AIV).  
 GN APOA4.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87089722; PubMed=3796595;  
 RA Williams S.C., Bruckheimer S.M., Lusis A.J., Lehoucq R.C.,  
 RA Kinniburgh A.J.;  
 RT "Mouse apolipoprotein A-IV gene: nucleotide sequence and induction by  
 a high-lipid diet."  
 RL Mol. Cell. Biol. 6:3807-3814(1986).  
 RN (2)  
 RP REVISIONS.  
 RA Kinniburgh A.J.;  
 RL Submitted (DEC-1986) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=C57BL/6J, and 129/J;  
 RX MEDLINE=91286309; PubMed=1648102;  
 RA Reue K., Lee T.H.;  
 RT "Genetic variation in mouse apolipoprotein A-IV due to insertion and  
 deletion in a region of tandem repeats."  
 RL J. Biol. Chem. 266:12715-12721(1991).  
 CC -I- FUNCTION: MAY HAVE A ROLE IN CHYLOMICRONS AND VLDL SECRETION AND  
 CATABOLISM. REQUIRED FOR EFFICIENT ACTIVATION OF LIPOPROTEIN  
 LIPASE BY APOC-II. POTENT ACTIVATOR OF LCAT. APOA-IV IS A MAJOR  
 COMPONENT OF HDL AND CHYLOMICRONS.  
 CC -I- SUBCELLULAR LOCATION: Extracellular.  
 CC -I- TISSUE SPECIFICITY: SECRETED IN PLASMA.  
 CC -I- DOMAIN: NINE OF THE THIRTEEN 22-AMINO ACID TANDDEM REPEATS (EACH  
 22-MER IS ACTUALLY A TANDDEM ARRAY OF TWO, A AND B, RELATED 11-  
 MERS) OCCURRING IN THIS SEQUENCE ARE PREDICTED TO BE HIGHLY ALPHA-  
 HELICAL, AND MANY OF THESE HELICES ARE AMPHIPATHIC. THEY MAY  
 THEREFORE SERVE AS LIPID-BINDING DOMAINS WITH LECTININ:CHOLESTEROL  
 ACUTRANSFERASE (LCAT) ACTIVATING ABILITIES.  
 CC -I- POLYMORPHISM: THERE IS A POLYMORPHISM WITHIN A SERIES OF IMPERFECT  
 REPEATS ENCODING THE SEQUENCE E-Q-[AV]-Q. INSERTIONS OR DELETIONS  
 OF 12 NUCLEOTIDES HAVE GIVEN RISE TO THREE FORMS CHARACTERIZED BY  
 THREE (129), FOUR (C57BL/6), OR FIVE (M.CASTANEUS) COPIES OF THE  
 REPEAT UNIT.  
 CC -I- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: M13966; AAA37253.1;  
 DR EMBL: M64249; AAA37215.1;  
 DR EMBL: M64248; AAA37214.1;  
 DR PIR: A25281; A25281.

DR MGD; MGI:88051; APOA4.  
 DR InterPro: IPR000074; Apolipoprotein.  
 DR Pfam: PF01442; Apolipoprotein; 2.  
 KW Plasma; Lipid transport; HDL; Chylomicron; Repeat; Signal;  
 KM Polymorphism.

FT	SIGNAL	1	20	395
FT	CHAIN	21	330	
FT	DOMAIN	33	330	
FT	REPEAT	33	54	
FT	REPEAT	60	81	
FT	REPEAT	82	103	
FT	REPEAT	115	136	
FT	REPEAT	137	158	
FT	REPEAT	159	180	
FT	REPEAT	181	202	
FT	REPEAT	203	224	
FT	REPEAT	225	246	
FT	REPEAT	247	268	
FT	REPEAT	269	286	
FT	REPEAT	287	308	
FT	REPEAT	309	330	
FT	DOMAIN	366	389	
FT	VARIANT	382	385	
FT	CONFLICT	15	15	
FT	CONFLICT	63	63	
FT	CONFLICT	207	207	
FT	CONFLICT	288	288	
FT	CONFLICT	294	295	
FT	CONFLICT	315	316	
SO	SEQUENCE	395 AA;	45029 MW;	5FE27D0236226257 CRC64;

Query Match 15.9%; Score 298.5; DB 1; Length 395;  
 Best Local Similarity 24.4%; Pred. No. 2.5e-11;  
 Matches 91; Conservative 85; Mismatches 164; Indels 33; Gaps 7;

OY 5 AAVLTALALISA-----FSATQARKGFMDYFSQTSQD-KGRVQDHOOKAREPATL-KD 58  
 DB 5 AAVLTALALVAITGRVAVTSDOAVANVWDYFTOLSNNAKEAVFOPTDVTQOLSTLFOD 64

OY 59 SLEODLNNMKNLEKRLPLSGSEAPRLPDPVGMROLOEELEVKARLOPYAAEAHELV 118  
 DB 65 KLDIASTYADGVANNKLVPFVVOVLSGLAKETPHVKEIKKELEDLNDMMHPANKYQTF 124

OY 119 GNNLEGLROOLKRYTMDIMQVALRVOELOEOLKVGCEDTKAOLGVD---EANAALQ 174  
 DB 125 GNNQKLOEHLKPYAVDLOQINTQOEMKLOLPYIQRTTIKEVNDLITSNMPLAT 184

OY 175 GLOSRVYVNHTRGRKELFHPYAESIASCIGRHOVLIHRSVAPHPASPARISSCVQVLSRK 234  
 DB 185 NIKDKRNKNNEHRLKCHLTTPRANLTKATIIIONLIDIKRSIALPIVGVQVQKINHOMEGIAFO 244

OY 235 LTKAKALARIQOONLDQLEELSRFACT-----GTGAGPDPOMLSEVRORLOAFR 289  
 DB 245 MKNNAEELQTKVSAKIDOLQKNLAPLVEDQSVKGTCTGLQSLSDLNQLOQVEFR 304

OY 290 QUDYVLOIAAFTRALIDETEPVOOQLAPPPGHSAPAFEPQOTDSKV---LSKLQARLD 346  
 DB 305 RYVEPGEEMKALVVOOLEFROOLGP-----NSGEVSHLSLEKSLRE 349

OY 347 LWEDITHSLHDG 359  
 DB 350 KVSFMSFTLEKKG 362

RESULT 7  
 APOA1\_HUMAN STANDARD; PRT; 267 AA.  
 AC P02647;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Apolipoprotein A-I precursor (Apo-AI).  
 GN APOA1.



- OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Furelestoml;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
NCBI\_TaxID=9606.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=64221405; PubMed=6328445;  
RA Sharpe C.R., Sidoti A., Shelley C.S., Lucero M.A., Shoulders C.C.,  
Baralle F.E.;  
RT "Human apolipoproteins AI, AII, CII and CIII. cDNA sequences and mRNA  
abundance.";  
RL Nucleic Acids Res. 12:3917-3932(1984).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=65026665; PubMed=6207999;  
RA Seilhammer J.J., Protter A.A., Frossard P., Levy-Wilson B.;  
RT "Isolation and DNA sequence of full-length cDNA and of the entire  
RT polymorphism in the apo AI gene.";  
RL DNA 3:309-317(1984).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=63220822; PubMed=6406984;  
RA Shoulders C.C., Kornblith A.R., Munro B.S., Baralle F.E.;  
RT "Gene structure of human apolipoprotein AI.";  
RL Nucleic Acids Res. 11:2827-2837(1983).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=63220772; PubMed=6304641;  
RA Cheung P., Chan L.;  
RT "Nucleotide sequence of cloned cDNA of human apolipoprotein A-I.";  
RL Nucleic Acids Res. 11:3703-3715(1983).  
RN [5]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=84119464; PubMed=6198645;  
RA Law S.W., Brewer H.B. Jr.;  
RT "Nucleotide sequence and the encoded amino acids of human  
RT apolipoprotein A-I mRNA.";  
RL Proc. Natl. Acad. Sci. U.S.A. 81:66-70(1984).  
RN [6]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86008382; PubMed=2995392;  
RA Law S.W., Brewer H.B. Jr.;  
RT "Tangier disease. The complete mRNA sequence encoding for  
RT preproapo-A-I.";  
RL J. Biol. Chem. 260:12810-12814(1985).  
RN [7]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=84016011; PubMed=6413973;  
RA Karathanasis S.K., Zannis V.I., Breslow J.L.;  
RT "Isolation and characterization of the human apolipoprotein A-I  
RT gene.";  
RL Proc. Natl. Acad. Sci. U.S.A. 80:6147-6151(1983).  
RN [8]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89377481; PubMed=2673706;  
RA Mogullevsky N., Roobol C., Loriau R., Guillaume J.P., Jacobs P.,  
RA Cravador A., Herzog A., Brouwers L., Scarso A., Gilles P.,  
RA Holmquist L., Carlson L.A., Bollen A.;  
RT "Production of human recombinant proapolipoprotein A-I in Escherichia  
RT coli: purification and biochemical characterization.";  
RL DNA 8:429-436(1989).  
RN [9]  
RP SEQUENCE FROM N.A. (VARIANT TANGIER).  
RX MEDLINE=88196137; PubMed=3129297;  
RA Medridis S.C., Ruiz-Opazo N., Hayden M., Nussbaum A.L., Breslow J.L.,  
RA Zannis V.I.;  
RT "Sequence and expression of Tangier apoA-I gene.";  
RL Eur. J. Biochem. 173:465-471(1988).  
RN [10]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skeletal muscle;  
RX Strausberg R.;
- RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.  
RN [11]  
RP SEQUENCE OF 118-267 FROM N.A.  
RX MEDLINE=83091059; PubMed=6294659;  
RA Breslow J.L., Ross D., McPherson J., Williams H.W., Kurnit D.,  
RA Nussbaum A.L., Karathanasis S.K., Zannis V.I.;  
RT "Isolation and characterization of cDNA clones for human  
RT apolipoprotein A-I.";  
RL Proc. Natl. Acad. Sci. U.S.A. 79:6861-6865(1982).  
RN [12]  
RP SEQUENCE OF 19-27.  
RX MEDLINE=83256553; PubMed=6409108;  
RA Brewer H.B. Jr., Fairwell T., Kay L., Meng M., Ronan K., Law S.,  
RA Light J.A.;  
RT "Human plasma proapoA-I: isolation and amino-terminal sequence.";  
RL Biochem. Biophys. Res. Commun. 113:626-632(1983).  
RN [13]  
RP SEQUENCE OF 25-267.  
RX MEDLINE=78123731; PubMed=204308;  
RA Brewer H.B. Jr., Fairwell T., Larue A., Ronan K., Houser A.,  
RA Bronzert T.J.;  
RT "The amino acid sequence of human APOA-I, an apolipoprotein isolated  
RT from high density lipoproteins.";  
RL Biochem. Biophys. Res. Commun. 80:623-630(1978).  
RN [14]  
RP SEQUENCE OF 25-267.  
RX MEDLINE=75133493; PubMed=164450;  
RA Baker H.N., Gotto A.M. Jr., Jackson R.L.;  
RT "The primary structure of human plasma high density apolipoprotein  
RT glutamine I (APOA-I). II. The amino acid sequence and alignment of  
RT cyanogen bromide fragments IV, III, and I.";  
RL J. Biol. Chem. 250:2725-2738(1975).  
RN [15]  
RP SEQUENCE OF 25-56.  
RX MEDLINE=88331387; PubMed=3047170;  
RA Yui Y., Aoyama T., Morishita H., Takahashi M., Takatsu Y., Kawai C.;  
RT "Serum prostacyclin stabilizing factor is identical to apolipoprotein  
RT A-I (Apo A-I). A novel function of Apo A-I.";  
RL J. Clin. Invest. 82:803-807(1988).  
RN [16]  
RP SEQUENCE OF 25-48.  
RX MEDLINE=89380318; PubMed=2506184;  
RA Manjunath P., Marcel Y.L., Uma J., Seidah N.G., Chretien M.,  
RA Chappelaine A.;  
RT "Apolipoprotein A-I binds to a family of bovine seminal plasma  
RT proteins.";  
RL J. Biol. Chem. 264:16853-16857(1989).  
RN [17]  
RP SEQUENCE OF 25-43.  
RX MEDLINE=88070603; PubMed=3120314;  
RA Pfiol R.P., Ordovas J.M., Rosenberg I., Schaeffer E.J.,  
RA Pereira M.E.A.;  
RT "Similarity of cruzin, an inhibitor of Trypanosoma cruzi  
RT neuraminidase, to high-density lipoprotein.";  
RL Science 238:1417-1419(1987).  
RN [18]  
RP SEQUENCE OF 25-42.  
RX MEDLINE=95203287; PubMed=7895732;  
RA Corbett J.M., Wheeler C.H., Baker C.S., Yacoub M.H., Dunn M.J.;  
RT "The human myocardial two-dimensional gel protein database: update  
RT 1994.";  
RL Electrophoresis 15:1459-1465(1994).  
RN [19]  
RP PALMITOYLATION.  
RX MEDLINE=86140194; PubMed=3005308;  
RA Hoeg J.M., Meng M.S., Ronan R., Fairwell T., Brewer H.B. Jr.;  
RT "Human apolipoprotein A-I. Post-translational modification by fatty  
RT acid acylation.";  
RL J. Biol. Chem. 261:3911-3914(1986).  
RN [20]  
RP PROCESSING.  
RX MEDLINE=83195100; PubMed=6405383;

RA Zanis V.I., Karathanasis S.K., Keutmann H.T., Goldberger G.,  
 RA Breslow J.L.,  
 RT "Intracellular and extracellular processing of human apolipoprotein  
 RT A-I: secreted apolipoprotein A-I isoprotein 2 is a propeptide.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:2574-2578(1983).  
 RN (21)  
 RN STRUCTURE BY NMR OF 190-209.  
 RX MEDLINE-96270776; PubMed-8664326;  
 RA Wang G., Treleaven W.D., Cushley R.J.;  
 RT "Conformation of human serum apolipoprotein A-I (166-185) in the  
 RT presence of sodium dodecyl sulfate or dodecylphosphocholine by 1H-NMR  
 RT and CD. Evidence for specific peptide-SDS interactions.";  
 RL Biochim. Biophys. Acta 1301:174-184(1996).  
 RN (22)  
 RN X-RAY CRYSTALLOGRAPHY (4.0 ANGSTROMS) OF 67-267.  
 RX MEDLINE-83109095; PubMed-6401735;  
 RA Weisgraber K.H., Rall S.C. Jr., Bersot T.P., Mahley R.W.,  
 RA Bortnick D.W., Rogers D.P., Ingler J.A., Brouillette C.G.;  
 RT "Crystal structure of truncated human apolipoprotein A-I suggests a  
 RT lipid-bound conformation.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:12291-12296(1997).  
 RN (23)  
 RN VARIANT MILANO.  
 RX MEDLINE-83109095; PubMed-6401735;  
 RA Weisgraber K.H., Rall S.C. Jr., Bersot T.P., Mahley R.W.,  
 RA Franceschini G., Sirtori C.R.;  
 RT "Apolipoprotein A-I Milano. Detection of normal A-I in affected  
 RT subjects and evidence for a cysteine for arginine substitution in the  
 RT variant A-I.";  
 RL J. Biol. Chem. 258:2508-2513(1983).  
 RN (24)  
 RN VARIANT TANGIER.  
 RX MEDLINE-83300108; PubMed-6412234;  
 RA Schmitz G., Assmann G., Rall S.C. Jr., Mahley R.W.;  
 RT "Tangier disease: defective recombination of a specific Tangier  
 RT apolipoprotein A-I isoform (pro-apo A-I) with high density  
 RT lipoproteins.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:6081-6085(1983).  
 RN (25)  
 RN VARIANT NORMAN.  
 RX MEDLINE-84289383; PubMed-6432779;  
 RA Rall S.C. Jr., Weisgraber K.H., Mahley R.W., Ogawa Y., Fielding C.J.,  
 RA Uleman G., Haas J., Steimel A., Menzel H.J., Assmann G.;  
 RT "Abnormal lecithin:cholesterol acyltransferase activation by a human  
 RT apolipoprotein A-I variant in which a single lysine residue is  
 RT deleted.";  
 RL J. Biol. Chem. 259:10063-10070(1984).  
 RN (26)

Query Match 10.7%; Score 200; DB 1; Length 267;  
 Best Local Simlarity 22.2%; Pred. No. 1.3e-05;  
 Matches 74; conservative 61; Mismatches 106; Indels 92; Gaps 10;

QY 5 AAVLTMAALISAFSAFOAKKQW-----YFSQSCIKGR--VEQIH 45  
 DB 3 AAVLTAVLEL--TGSQAH--FWQDEPQSWDRVKDLATVYDVYKSGRDYVSQFE 58  
 QY 46 QOKMARE-PATLKDSLIEDIINNKKFLKRLPLSGSEAPRLPODPYQMRLOEELBEVK 104  
 DB 59 GSAIGQAINIKLNDNDVSTSTSKLRQGLPTQOFEMNLKRYTEGLKQEMSKDLEEVK 118  
 QY 105 ARIQPYMAENHELVGNNEGLRQOLKPYTMDLMEQVALRVELOEDLRVVGEDTKAOLG 164  
 DB 119 AKQAPYLDFOKKQWQEMELYLROKVERPLRAELOEGAROKLHELOEKLSPIGEMR---- 173  
 QY 165 GVDEAMALLQGLQSVVNHGRRKLLFHYAEESLVGICRHVQELRSVAPRPARL 224  
 DB 174 --DRAA-----HVDALTHLAPYSDELRL 198  
 QY 225 SRGVYVSRKTLKAKALHARLDQNLQDLREELSRFAFGTGEAGCPDQMLSEEVOR 284  
 DB 199 AARLEALKENGARLAEVYAKAKATEHLSTLSEKKKRAL-----EDLRQG 241  
 QY 285 ----LQAFRODTYLOIAATFTRALDQTEVEVQO 313

DB 242 LLPVLESF-----VSFLSALEETKKNLTQ 267  
 ID APAL\_MACFA STANDARD; PRT: 267 AA.  
 AC P15568; P17929;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Apolipoprotein A-I precursor (Apo-AI).  
 GN APOA1.  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), and  
 OS Papio hamadryas (Hamadryas baboon).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecinae; Macaca.  
 OX NCBI\_TaxID:9541, 9557;  
 RN (1)  
 RN SEQUENCE FROM N.A.  
 KC SPECIES-M.fascicularis;  
 RX MEDLINE-87191989; PubMed-3106152;  
 RA Polites H.G., Melchior G.W., Castile C.K., Marolli K.R.;  
 RT "The primary structure of cynomolgus monkey apolipoprotein A-I  
 RT deduced from the cDNA sequence: comparison to the human sequence.";  
 RL Gene 49:103-110(1986).  
 RN (2)  
 RN SEQUENCE FROM N.A.  
 KC SPECIES-M.fascicularis;  
 RX MEDLINE-92305062; PubMed-1610902;  
 RA Murray R.W., Marolli K.R.;  
 RT "Nucleotide sequence of the cynomolgus monkey apolipoprotein A-I gene  
 RT and corresponding flanking regions.";  
 RL Biochim. Biophys. Acta 1131:207-210(1992).  
 RN (3)  
 RN SEQUENCE OF 25-48.  
 KC SPECIES-M.fascicularis;  
 RX MEDLINE-87185451; PubMed-3105581;  
 RA Herbert P.N., Hausserman L.L., Lynch K.M., Sartelli A.L.,  
 RA Kantor M.A., Nicolosi R.J., Shulman R.S.;  
 RT "Homologues of the human C and A apolipoproteins in the Macaca  
 RT fascicularis (cynomolgus) monkey.";  
 RL Biochemistry 26:1457-1463(1987).  
 RN (4)  
 RN SEQUENCE FROM N.A.  
 KC SPECIES-P.hamadryas; TISSUE=Liver;  
 RX MEDLINE-89232739; PubMed-2907746;  
 RA Blixson J.E., Borenstein S., Cox L.A., Rainwater D.L., Vandeberg J.L.;  
 RT "The baboon gene for apolipoprotein A-I: characterization of a cDNA  
 RT clone and identification of DNA polymorphisms for genetic studies of  
 RT cholesterol metabolism.";  
 RL Gene 74:483-490(1988).  
 RN (5)  
 RN SEQUENCE OF 1-10 FROM N.A.  
 RC SPECIES-M.fascicularis;  
 RA Sorci-Thomas M.;  
 RT Submitted (Oct-1991) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: APOA-1 PARTICIPATES IN THE REVERSE TRANSPORT OF  
 CC CHOLESTEROL FROM TISSUES TO THE LIVER FOR EXCRETION BY PROMOTING  
 CC CHOLESTEROL EFFLUX FROM TISSUES AND BY ACTING AS A COFACTOR FOR  
 CC THE LECTININ CHOLESTEROL ACYLTRANSFERASE (LCAT).  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- TISSUE SPECIFICITY: MAJOR PROTEIN OF PLASMA HDL, ALSO FOUND IN  
 CC CHYLOMICRONS.  
 CC -1- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to [license@db-sib.ch](mailto:license@db-sib.ch).

CC EMBL: M15411; AAA36834.1; -  
 DR EMBL: M83242; AAA36832.1; -  
 DR EMBL: M35634; AAA35380.1; -  
 DR EMBL: M69223; AAA36831.1; -  
 DR PIR: A26529; A26529.  
 DR PIR: A26627; A26627.  
 DR PIR: JS0079; JS0079.  
 DR PIR: S23135; S23135.  
 DR HSSP: P02647; IGW4.  
 DR InterPro: IPR000074; Apolipoprotein.  
 DR Pfam: Pf01442; Apolipoprotein; 1.  
 KW Plasma; Lipid transport; HDL; Cholesterol metabolism; Repeat; Signal.  
 FT STGNL 1 18  
 FT PROPEP 19 24  
 FT CHAIN 25 267  
 FT DOMAIN 68 267  
 FT REPEAT 68 89  
 FT REPEAT 90 111  
 FT REPEAT 112 122  
 FT REPEAT 123 144  
 FT REPEAT 145 166  
 FT REPEAT 167 188  
 FT REPEAT 189 210  
 FT REPEAT 211 232  
 FT REPEAT 233 243  
 FT REPEAT 244 267  
 FT CONFLICT 13 13  
 SQ SHOUNCE 267 AA; 30735 MW; 869955C0408BE21 CRC64;  
 L -> P (IN REF. 1).  
 Query Match 10.3%; Score 193; DB 1; Length 267;  
 Best Local Similarity 22.2%; Pred. No. 3.5e-05;  
 Matches 74; Conservative 61; Mismatches 106; Indels 92; Gaps 10;

CC EMBL: M15908; CAA34024.1; -  
 DR EMBL: X06658; CAA29857.1; -  
 DR EMBL: X06659; CAA29858.1; -  
 DR PIR: S06064; LPRB12.  
 DR PIR: S00230; LPRB18.  
 DR PIR: A24998; A24998.  
 DR HSSP: P02647; IAV1.  
 DR InterPro: IPR000074; Apolipoprotein.  
 DR Pfam: Pf01442; Apolipoprotein; 1.  
 KW Plasma; Lipid transport; HDL; Cholesterol metabolism; Repeat; Signal.  
 FT STGNL 1 18  
 FT PROPEP 19 24  
 FT CHAIN 25 266  
 FT DOMAIN 67 266  
 FT REPEAT 67 88  
 FT REPEAT 89 110  
 FT REPEAT 111 121  
 FT REPEAT 112 143  
 FT REPEAT 144 165  
 FT REPEAT 166 187  
 FT REPEAT 188 209  
 FT REPEAT 210 231  
 FT REPEAT 232 242  
 FT REPEAT 243 266  
 FT CONFLICT 18 18  
 FT CONFLICT 44 44  
 FT CONFLICT 45 45  
 FT CONFLICT 107 107  
 FT CONFLICT 123 123  
 FT CONFLICT 147 147  
 FT CONFLICT 150 150  
 FT CONFLICT 191 191  
 FT CONFLICT 195 195  
 FT CONFLICT 211 211

APOLIPROTEIN A-I.  
 10 X APPROXIMATE TANDEM REPEATS.  
 1.  
 2.  
 3 (HALF-LENGTH).  
 4.  
 5.  
 6.  
 7.  
 8.  
 9 (HALF-LENGTH).  
 10.  
 A -> R (IN REF. 2; CAA29858).  
 MISSING (IN REF. 3).  
 V -> I (IN REF. 2; CAA29858).  
 Y -> F (IN REF. 3).  
 Y -> V (IN REF. 2; CAA29857).  
 A -> G (IN REF. 2; CAA29858 AND 3).  
 R -> G (IN REF. 2; CAA29858/CAA29857).  
 N -> Q (IN REF. 3).  
 MISSING (IN REF. 2; CAA29858/CAA29857).  
 S -> K (IN REF. 2; CAA29858/CAA29857).

CC EMBL: M15908; CAA34024.1; -  
 DR EMBL: X06658; CAA29857.1; -  
 DR EMBL: X06659; CAA29858.1; -  
 DR PIR: S06064; LPRB12.  
 DR PIR: S00230; LPRB18.  
 DR PIR: A24998; A24998.  
 DR HSSP: P02647; IAV1.  
 DR InterPro: IPR000074; Apolipoprotein.  
 DR Pfam: Pf01442; Apolipoprotein; 1.  
 KW Plasma; Lipid transport; HDL; Cholesterol metabolism; Repeat; Signal.  
 FT STGNL 1 18  
 FT PROPEP 19 24  
 FT CHAIN 25 266  
 FT DOMAIN 67 266  
 FT REPEAT 67 88  
 FT REPEAT 89 110  
 FT REPEAT 111 121  
 FT REPEAT 112 143  
 FT REPEAT 144 165  
 FT REPEAT 166 187  
 FT REPEAT 188 209  
 FT REPEAT 210 231  
 FT REPEAT 232 242  
 FT REPEAT 243 266  
 FT CONFLICT 18 18  
 FT CONFLICT 44 44  
 FT CONFLICT 45 45  
 FT CONFLICT 107 107  
 FT CONFLICT 123 123  
 FT CONFLICT 147 147  
 FT CONFLICT 150 150  
 FT CONFLICT 191 191  
 FT CONFLICT 195 195  
 FT CONFLICT 211 211

APOLIPROTEIN A-I.  
 10 X APPROXIMATE TANDEM REPEATS.  
 1.  
 2.  
 3 (HALF-LENGTH).  
 4.  
 5.  
 6.  
 7.  
 8.  
 9 (HALF-LENGTH).  
 10.  
 A -> R (IN REF. 2; CAA29858).  
 MISSING (IN REF. 3).  
 V -> I (IN REF. 2; CAA29858).  
 Y -> F (IN REF. 3).  
 Y -> V (IN REF. 2; CAA29857).  
 A -> G (IN REF. 2; CAA29858 AND 3).  
 R -> G (IN REF. 2; CAA29858/CAA29857).  
 N -> Q (IN REF. 3).  
 MISSING (IN REF. 2; CAA29858/CAA29857).  
 S -> K (IN REF. 2; CAA29858/CAA29857).

CC EMBL: M15908; CAA34024.1; -  
 DR EMBL: X06658; CAA29857.1; -  
 DR EMBL: X06659; CAA29858.1; -  
 DR PIR: S06064; LPRB12.  
 DR PIR: S00230; LPRB18.  
 DR PIR: A24998; A24998.  
 DR HSSP: P02647; IAV1.  
 DR InterPro: IPR000074; Apolipoprotein.  
 DR Pfam: Pf01442; Apolipoprotein; 1.  
 KW Plasma; Lipid transport; HDL; Cholesterol metabolism; Repeat; Signal.  
 FT STGNL 1 18  
 FT PROPEP 19 24  
 FT CHAIN 25 266  
 FT DOMAIN 67 266  
 FT REPEAT 67 88  
 FT REPEAT 89 110  
 FT REPEAT 111 121  
 FT REPEAT 112 143  
 FT REPEAT 144 165  
 FT REPEAT 166 187  
 FT REPEAT 188 209  
 FT REPEAT 210 231  
 FT REPEAT 232 242  
 FT REPEAT 243 266  
 FT CONFLICT 18 18  
 FT CONFLICT 44 44  
 FT CONFLICT 45 45  
 FT CONFLICT 107 107  
 FT CONFLICT 123 123  
 FT CONFLICT 147 147  
 FT CONFLICT 150 150  
 FT CONFLICT 191 191  
 FT CONFLICT 195 195  
 FT CONFLICT 211 211

APOLIPROTEIN A-I.  
 10 X APPROXIMATE TANDEM REPEATS.  
 1.  
 2.  
 3 (HALF-LENGTH).  
 4.  
 5.  
 6.  
 7.  
 8.  
 9 (HALF-LENGTH).  
 10.  
 A -> R (IN REF. 2; CAA29858).  
 MISSING (IN REF. 3).  
 V -> I (IN REF. 2; CAA29858).  
 Y -> F (IN REF. 3).  
 Y -> V (IN REF. 2; CAA29857).  
 A -> G (IN REF. 2; CAA29858 AND 3).  
 R -> G (IN REF. 2; CAA29858/CAA29857).  
 N -> Q (IN REF. 3).  
 MISSING (IN REF. 2; CAA29858/CAA29857).  
 S -> K (IN REF. 2; CAA29858/CAA29857).

FT CONFLICT 255 256 VL -> LV (IN REF. 3)  
 FT CONFLICT 256 256 L -> V (IN REF. 2; CAA29858/CAA29857).  
 SO SOURCE 256 AA: 30591 MW: 0FF60D386497C7D2 CRC64:  
 Query Match 9.9%; Score 185; DB 1; Length 266;  
 Best Local Similarity 22.4%; Pred. No. 0.0001;  
 Matches 61; Conservative 61; Mismatches 122; Indels 28; Gaps 6;

QY 5 AAVLTWALALISAFSAIOAR-----KCFWD-----YFSQSGDKCR--VBOIHOO 47  
 Db 3 AVVLTWALVFL--TGSOARHFWQHDDEPRSSWDKIKDFATVYVDVTKDSGREYVAQFEAS 59  
 QY 48 KWARE-PATIKDSLEQDINNMMKFEKLRLPLSGSEAPRLPODPVGMRRLOLELEEVKAR 106  
 Db 60 AFGKOLNKLIDNMDSLSSTVSKLOEQLQVDFWDMLEKTEGILKEPMKDIQEVKOK 119  
 QY 107 IQPYAAEAHELIVGNNIFGILKQQLKPYTMDLMEQVALRVQLOEQLRVYCEIDPKAOLLCGV 166  
 Db 120 VQPYLDEFOKKWQEVERYRQKVEPLGALRESARQKLTLEQEKSLPLAEELR----- 172  
 QY 167 DEAMALLQGLQSRVYVHTORKEFLFHPYAESLVSGIGRNVOLHRSVADHPASPAPRLSR 226  
 Db 173 DSARTHTVPLTKLAPYSNEUQRLAARLESIKKCGASLAEYQAKARKVHLSVLSKARP 232  
 QY 227 CVQVLSRKTLTKAKALHARIQONLDQLEELS 258  
 Db 233 ALEDRLQGLIPVLESFKASVQNVLDHATKIKLN 264

RESULT 10  
 ID APAL CHICK STANDARD: PRT: 264 AA.  
 AC P08250:  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Apolipoprotein A-I precursor (Apo-AI).  
 GN APOA1.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mollusca; Mollusca; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianidae; Gallus.  
 OC NCBI\_TaxID:9031:  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=88049703; PubMed=3118875;  
 BY Bynes L., Luo C.-C., Li W.-H., Yang C.-Y., Chan L.;  
 RT "Chicken apolipoprotein A-I: cDNA sequence, tissue expression and evolution.";  
 RL Biochem. Biophys. Res. Commun. 148:485-492(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=88152500; PubMed=3126099;  
 BY Ferrari S., Tarugi P., Drusiani E., Calandra S., Fregni M.;  
 RT "The complete sequence of chicken apolipoprotein A-I mRNA and its expression in the developing chick.";  
 RL Gene 60:39-46(1987).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=87223301; PubMed=3108248;  
 BY Rajavashisth T.B., Dawson P.A., Williams D.L., Shackelford J.E.,  
 RA Leberer H., Lusis A.J.;  
 RT "Structure, evolution, and regulation of chicken apolipoprotein A-I.";  
 RL J. Biol. Chem. 262:7056-7065(1987).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=92381402; PubMed=1512510;  
 BY Lamon-Fava S., Sastri R., Ferrari S., Rajavashisth T.B.,  
 RA Lusis A.J., Karathanasis S.K.;  
 RT "Evolutionary distinct mechanisms regulate apolipoprotein A-I gene expression: differences between avian and mammalian apoA-I gene transcription control regions.";  
 RL J. Lipid Res. 33:831-842(1992).

RN [5]  
 RP SEQUENCE OF 25-44.  
 RA MEDLINE=83213468; PubMed=6406496;  
 RA Shackelford J.E., Leberer H.G.;  
 RT "Synthesis and secretion of apolipoprotein A1 by chick breast muscle.";  
 RL J. Biol. Chem. 258:7175-7180(1983).  
 CC - FUNCTION: APOA-1 PARTICIPATES IN THE REVERSE TRANSPORT OF  
 CC CHOLESTEROL FROM TISSUES TO THE LIVER FOR EXCRETION BY PROMOTING  
 CC THE EFFLUX OF TISSUE CHOLESTEROL AND TRIGLYCERIDES AS A COFACTOR FOR  
 CC THE LIPIDIN CHOLESTEROL ACYLTRANSFERASE (LCAT).  
 CC - SUBCELLULAR LOCATION: Extracellular.  
 CC - TISSUE SPECIFICITY: MAJOR PROTEIN OF PLASMA HDL, ALSO FOUND IN  
 CC CHYLOMICRONS.  
 CC - SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOH FAMILY.  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb.ch](mailto:license@isb.ch)).

DR EMBL: M17961; AAA48593.1;  
 DR EMBL: M18746; AAA48594.1;  
 DR EMBL: M25559; AAA48592.1;  
 DR EMBL: M96012; AAA48597.1;  
 DR PIR: S01453; LPPHAI.  
 DR PIR: JH0471; JH0471.  
 DR HSSP: P02647; LGW4.  
 DR InterPro: IPR000074; Apolipoprotein.  
 DR Pfam: PF01442; Apolipoprotein.  
 DR KMW Plasma; Lipid transport; HDL; Cholesterol metabolism; Repeat; Signal.  
 FT SIGNAL 1 18  
 FT PROPEP 19 24  
 FT DOMAIN 25 264  
 FT REPEAT 67 264  
 FT REPEAT 67 88  
 FT REPEAT 89 110  
 FT REPEAT 111 121  
 FT REPEAT 122 143  
 FT REPEAT 144 165  
 FT REPEAT 166 187  
 FT REPEAT 188 209  
 FT REPEAT 210 231  
 FT REPEAT 232 242  
 FT REPEAT 243 264  
 FT CONFLICT 16 16  
 FT CONFLICT 148 148  
 SO SOURCE 264 AA: 30680 MW: 968320E81E2AC5C2 CRC64:

Query Match 9.7%; Score 183; DB 1; Length 264;  
 Best Local Similarity 21.1%; Pred. No. 0.00014;  
 Matches 70; Conservative 62; Mismatches 113; Indels 86; Gaps 9;

QY 4 MAVLTWALALISAFSAIOAR-----DYFSQT--SCDKGREVOIH 45  
 Db 1 MRCVLY-FLAVL-FLTGQAR-SFWQHDDEPQVLDRIKDMVDVLETVKASGKDAIAQFE 57  
 QY 46 OOKMARE-PATIKDSLEQDINNMMKFEKLRLPLSGSEAPRLPODPVGMRRLOLELEEVK 104  
 Db 58 SSAVKQKDLKADLNDLTLSSAAAKLRQDMAPYKREVRMMKLTDEALRAELTLDLEVK 117  
 QY 105 ARLQPYAAEAHELIVGNNIFGILKQQLKPYTMDLMEQVALRVQLOEQLRVYCEIDPKAOLLCGV 164  
 Db 118 EKIRFELDFQSAKWTEDEQYRQRLTPVAQGLKELTKVQELMKAQLTPVAEARDRLRG 177  
 QY 165 GVDEAMALLQGLQSRVYVHTORKEFLFHPYAESLVSGIGRNVOLHRSVADHPASPAPRL 224  
 Db 178 -----HVEELRNKLAPYSDELROKL 197  
 QY 225 SRCVQVLSRKTLTKAKALHARIQONLDQLEELSRAFAGTOTEGGAPDPQKMLSEEVKOR 284

DB 198 SOKBEIRENGIAPASEVQAKVMQSLNLRKMT-----PLVGEFRER 240  
 OY 285 LOAFRDTYQIAAFTRAIDQETEEVQOOLA 315  
 DB 241 LIPYAENLNKRLISF-----LDRLOKSA 264

## RESULT 11

API2\_ONCMY STANDARD; PRT: 262 AA.  
 ID API2\_ONCMY  
 AC 057524;  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Apolipoprotein A-I-2 precursor (APOA-I-2).  
 OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.  
 OX NCBI\_TaxID=8022;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Shasta; Tissue=Liver;  
 RX MEDLINE=92235558; PubMed=1569377;  
 RA Delcuve G.P., Sun J.M., Davie J.R.;  
 RT "Expression of rainbow trout apolipoprotein A-I genes in liver and  
 hepatocellular carcinoma."  
 RL J. Lipid Res. 33:251-262(1992).

CC -1- FUNCTION: APOA-1 PARTICIPATES IN THE REVERSE TRANSPORT OF  
 CHOLESTEROL FROM TISSUES TO THE LIVER FOR EXCRETION BY PROMOTING  
 CHOLESTEROL EFFLUX FROM TISSUES AND BY ACTING AS A COFACTOR FOR  
 THE LECTIN CHOLESTEROL ACYLTRANSFERASE (LCAT) (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Extracellular (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AF042219; AAB96973.1;  
 DR InterPro: IPR000074; Apolipoprotein.  
 DR Pfam: PF01442; Apolipoprotein.1.  
 KM Plasma; Lipid transport; HDL; Cholesterol metabolism; Repeat; Signal.  
 FT SIGNAL 1 18  
 FT PROPEP 19 23 POTENTIAL.  
 FT CHAIN 24 262 APOLOPOPROTEIN A-I-2.  
 FT DOMAIN 32 63 3 X APPROXIMATE TANDEM REPEATS.  
 FT DOMAIN 64 262 10 X APPROXIMATE TANDEM REPEATS.  
 FT REPEAT 67 85 1.  
 FT REPEAT 87 107 2.  
 FT REPEAT 108 118 3 (HALF-LENGTH).  
 FT REPEAT 119 140 4.  
 FT REPEAT 141 162 5.  
 FT REPEAT 163 184 6.  
 FT REPEAT 185 206 7.  
 FT REPEAT 207 228 8.  
 FT REPEAT 229 239 9 (HALF-LENGTH).  
 FT REPEAT 240 262 10.  
 SQ SEQUENCE 262 AA; 29680 MW; 234213D38B511A64 CRC64;

Query Match 9.58; Score 179; DB 1; Length 262;  
 Best Local Similarity 22.58; Pred. No. 0.00024;  
 Matches 60; Conservative 62; Mismatches 107; Indels 38; Gaps 7;

OY 8 LTNALLLSFSATQA-----RKGFMDYFSQTSQSGKRGEOHQKMAR-- 51  
 DB 4 LALALITLLA-AATQAVPMQADAPSLQLEHYKVAAMETWAO-----VKETAQNSIDRLD 55

OY 52 ---EPATLKDSLBDLNNNNKFL---EKLRLPSGSEAPRLPDDPYGMRQLOEELREYKA 105  
 DB 56 DTEYKEVYKVSQSLDNLQIAQTAQSAESLAPSEAIQVLETAVAERAKDVEELRS 115  
 OY 106 RIQPYMAEHAHELWGNLEGLROOLKPYTMDLMEQVALRVOELOEDLRVVGEDTKQLLGG 165  
 DB 116 QLEPRRAELKEVLDKHIDYEYRKLEPLDKDIVEQRTLEAFRAVLEPVEELMRKAVSAN 175  
 OY 166 VDEAMA---LIQLOSRVYHHTGRFKEFLPIYAESLVSGIRHVOELHRSVAPHPASP 221  
 DB 176 VEETRAKLMPYIEYVRAKILTERLEELRLASPYAEYKEQWAKAVGEVREKVPPLTDFK 235  
 OY 222 ARLSRCVOYLAKLTLKAKALHARIQO 248  
 DB 236 GQLGPAAEQAKEL---MALYETISQ 258

## RESULT 12

APAL\_BOVIN STANDARD; PRT: 265 AA.  
 ID APAL\_BOVIN  
 AC P15497;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Apolipoprotein A-I precursor (Apo-AI).  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 OX [1]

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90348478; PubMed=2117227;  
 RA O'Huigin C., Chan L., Li W.H.;  
 RT "Cloning and sequencing of bovine apolipoprotein A-I cDNA and  
 molecular evolution of apolipoproteins A-I and B-100."  
 RL Mol. Biol. Evol. 7:327-339(1990).

RN [2]  
 RP SEQUENCE OF 19-265.  
 RX MEDLINE=92153895; PubMed=1739745;  
 RA Sparrow D.A., Lee B.R., Laplaud M.P., Auboitron S., Beauchart D.,  
 RA Chapman J.M., Goto A.M., Jr., Yang C.Y., Sparrow J.T.;  
 RT "Plasma lipid transport in the preterm calf, Bos spp. primary  
 structure of bovine apolipoprotein A-I."  
 RL Biochim. Biophys. Acta 1123:145-150(1992).

RN [3]  
 RP SEQUENCE OF 25-70.  
 RX MEDLINE=90147795; PubMed=2105728;  
 RA Auboitron S., Sparrow D.A., Beauchat L., Beauchart D., Sparrow J.T.,  
 RA Laplaud M.P., Chapman J.M.;  
 RT "Characterization and amino-terminal sequence of apolipoprotein AI  
 from plasma high density lipoproteins in the preterm calf, Bos  
 spp."  
 RL Biochem. Biophys. Res. Commun. 166:833-839(1990).

CC -1- FUNCTION: APOA-1 PARTICIPATES IN THE REVERSE TRANSPORT OF  
 CHOLESTEROL FROM TISSUES TO THE LIVER FOR EXCRETION BY PROMOTING  
 CHOLESTEROL EFFLUX FROM TISSUES AND BY ACTING AS A COFACTOR FOR  
 THE LECTIN CHOLESTEROL ACYLTRANSFERASE (LCAT).  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- TISSUE SPECIFICITY: MAJOR PROTEIN OF PLASMA HDL, ALSO FOUND IN  
 CHYLOMICRONS.  
 CC -1- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: M35870; AAA30381.1; -

DR PIR: A34649; A34649.  
 DR HSSP: P02647; 1AV1.  
 DR InterPro: IPR000074; Apolipoprotein.  
 DR Pfam: PF01442; Apolipoprotein; 1.  
 KW Plasma; Lipid transport; HDL; Cholesterol metabolism; Repeat; Signal.  
 FT SIGNAL 1 18  
 FT PROPEP 19 24  
 FT CHAIN 25 265  
 FT DOMAIN 67 265  
 FT REPEAT 67 88  
 FT REPEAT 89 110  
 FT REPEAT 111 121  
 FT REPEAT 122 143  
 FT REPEAT 144 165  
 FT REPEAT 166 187  
 FT REPEAT 188 209  
 FT REPEAT 210 230  
 FT REPEAT 231 241  
 FT REPEAT 242 265  
 FT CONFLICT 185 186  
 FT SEQUENCE 265 AA; 30276 MW; 06A2681EA2ABA50F CRC64;  
 Query Match 9.4%; Score 176; DB 1; Length 265;  
 Best Local Similarity 22.0%; Pred. No. 0.00036;  
 Matches 74; Conservative 58; Mismatches 103; Indels 102; Gaps 10;  
 OY 5 AAVLTWALLLSAFSATOAR-----KGFMDYFSQTSQDKGR--VEQIHQ 47  
 DB 3 AAVLTAVLFL--TSGQARHFWMQDDPOSSMDRVDKDFATVYEAVIDSGRDYVAQFEAS 59  
 OY 48 KMARE-PATLKDSLEQDLNNMKNKFLKRLPSGSEARILPDDPYGMKROLOEELFEKKAR 106  
 DB 60 ALGQNLKLLDMWDLTASTLSKVRQELGVTOEFMDNLEKETSLOEMNKDLEEVKQK 119  
 OY 107 LOPYMAEHLVGMNLEGLRQOLKPYTMDMEQVALRVQEOLEQRLRVGSEDTKAQLLGGV 166  
 DB 120 VOPYLDEFQKKHVEVEIYRQKVAPLCEFRREGARQKVOELQDL----- 164  
 OY 167 DEAMALLQGLQSRVYVHHTGFKELFHYAESLVSGIRHVOELHRSVAPHPASPARLSR 226  
 DB 165 -----SPLAQLDRARAHYETLRQQLAPYS----- 190  
 OY 227 CVOVLSKRLTKAKKAL-----HARIQONLDQRELSNAFACITGECAGPDPM 276  
 DB 191 --DDLRLRLARLEALKEGGSLAEYHAKKASEQKALGK-----AKP----- 231  
 OY 277 LSEEVQRLOAFRODTYLOIAFTRAIDQETEEVQOQ 313  
 DB 232 VLEDRLQGLLPVLSLKVSIIA--AIDVASKKIMAQ 265  
 RESULT 13  
 APAL-CANFA STANDARD; PRT; 266 AA.  
 AC P02648;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Apolipoprotein A-I precursor (Apo-AI).  
 CN APOA1.  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OC NCBI\_TaxID=9615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RA MEDLINE=90132271; PubMed=2515239;  
 RA Luo C.-C., Li W.-H., Chan L.;  
 RT "Structure and expression of dog apolipoprotein A-I, E, and C-I  
 RT MNAs: implications for the evolution and functional constraints of  
 RT apolipoprotein structure";  
 RL J. Lipid Res. 30:1735-1746(1989).

RN [2]  
 RP SEQUENCE OF 25-266.  
 RX MEDLINE=82142425; PubMed=6801039;  
 RA Chung H., Randolph A., Reardon I., Heinrichson R.L.;  
 RT "The covalent structure of apolipoprotein A-I from canine high  
 RT density lipoproteins";  
 RL J. Biol. Chem. 257:2961-2967(1982).  
 RN [3]  
 RP SEQUENCE OF 25-57 AND 262-265.  
 RX MEDLINE=76210910; PubMed=179887;  
 RA Nakai T., Whayne T.F., Tang J.;  
 RT "The amino- and carboxyl-terminal sequences of canine apolipoprotein  
 RT A-I";  
 RL FEBS Lett. 64:409-411(1976).  
 RN [4]  
 RP SEQUENCE OF 25-37.  
 RC TISSUE=Heart;  
 RX MEDLINE=9816340; PubMed=9504812;  
 RA Dunn M.J., Corbett J.M., Wheeler C.H.;  
 RT "HSC-2DPAGE and the two-dimensional gel electrophoresis database of  
 RT dog heart proteins";  
 RL Electrophoresis 18:2795-2802(1997).  
 CC -1- FUNCTION: APOA-1 PARTICIPATES IN THE REVERSE TRANSPORT OF  
 CC CHOLESTEROL FROM TISSUES TO THE LIVER FOR EXCRETION BY PROMOTING  
 CC CHOLESTEROL EFFLUX FROM TISSUES AND BY ACTING AS A COPACITOR FOR  
 CC THE LECTITHIN CHOLESTEROL ACYLTRANSFERASE (LCAT).  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- TISSUE SPECIFICITY: MAJOR PROTEIN OF PLASMA HDL, ALSO FOUND IN  
 CC CHYLOMICRONS. SYNTHESIZED IN THE LIVER AND SMALL INTESTINE.  
 CC -1- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.  
 DR PIR: A03092; LPDCA1.  
 DR HSSP: P02647; 1AV1.  
 DR HSC-2DPAGE; P02648; DOG.  
 DR InterPro: IPR000074; Apolipoprotein.  
 DR Pfam: PF01442; Apolipoprotein; 1.  
 KW Plasma; Lipid transport; HDL; Cholesterol metabolism; Repeat; Signal.  
 FT SIGNAL 1 18  
 FT PROPEP 19 24  
 FT CHAIN 25 266  
 FT DOMAIN 67 266  
 FT REPEAT 67 88  
 FT REPEAT 89 110  
 FT REPEAT 111 121  
 FT REPEAT 122 143  
 FT REPEAT 144 165  
 FT REPEAT 166 187  
 FT REPEAT 188 209  
 FT REPEAT 210 231  
 FT REPEAT 232 242  
 FT REPEAT 243 266  
 FT CONFLICT 168 168  
 FT CONFLICT 202 202  
 FT CONFLICT 235 235  
 FT CONFLICT 264 266  
 FT SEQUENCE 266 AA; 30196 MW; A3202620C28A869D CRC64;  
 Query Match 9.4%; Score 175.5; DB 1; Length 266;  
 Best Local Similarity 22.0%; Pred. No. 0.00039;  
 Matches 72; Conservative 53; Mismatches 120; Indels 83; Gaps 8;  
 OY 5 AAVLTWALLLSAFSATOARKCFMD-----YFSQTSQDKGR--VEQIHQ 46  
 DB 3 AAVLTAVLFL--TSGQARH-FMQDDPOSSMDRVDKDLATYVADAVKSGRDYVAQFEA 58  
 OY 47 QKMARE-PATLKDSLEQDLNNMKNKFLKRLPSGSEARILPDDPYGMKROLOEELFEVKA 105  
 DB 59 SALGQNLKLLDMWDLTASTLSKVRQELGVTOEFMDNLEKETSLOEMNKDLEEVKQ 118  
 OY 106 LOPYMAEHLVGMNLEGLRQOLKPYTMDMEQVALRVQEOLEQRLRVGSEDTKAQLLGG 165  
 DB 119 KOPYLDDQKKHVEVEIYRQKVAPLGSELREGARQKVOELQDL----- 164  
 OY 166 VDEAMALLQGLQSRVYVHHTGFKELFHYAESLVSGIRHVOELHRSVAPHPASPARLS 225

DB 165 -----SPLAEELDRARHTVDALRALAPYSDDLRELA 198  
 OY 226 RCYQVIVSRKLTAKALHARIQONLDOLKEHLSRAFACTGTEGACHPDQMLSEVQRRL 285  
 DB 199 ARLEALEKGGASLAETHARASEQLSALGEXARPAL-----EDLRQGL 241  
 OY 286 QAFRODTYLOIAAFTRAIDQETEEVQOQ 313  
 DB 242 LPVLESFVKVSLA---AIDEXTKLMAQ 266

RESULT 14  
 APAL\_ANAPL STANDARD; PRT: 264 AA.  
 ID APAL\_ANAPL STANDARD; PRT: 264 AA.  
 AC 042336;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Apolipoprotein A-I precursor (Apo-AI).  
 GN APOA1.  
 OS Anas platyrhynchos (Domestic duck).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Anseriformes; Anellidae; Anas.  
 OX NCBI\_TaxID=8839;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Pekin breed; TISSUE=Liver;  
 RA Chen B., Lu X., Wang K., Xue H., Zeng W., Zhu D., Chi L.;  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: APOA-1 PARTICIPATES IN THE REVERSE TRANSPORT OF  
 CC CHOLESTEROL FROM TISSUES TO THE LIVER FOR EXCRETION BY PROMOTING  
 CC THE LECTININ CHOLESTEROL ACYLTRANSFERASE (LCAT).  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- TISSUE SPECIFICITY: MAJOR PROTEIN OF PLASMA HDL, ALSO FOUND IN  
 CC CHYLOMICRONS.  
 CC -1- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: U86131; AAB64381.1;  
 CC HSSP: P02647; IGM4.  
 DR Interpro: IIR000074; Apolipoprotein.  
 DR Pfam: PF01472; Apolipoprotein\_1.  
 KW Plasma; Lipid transport; HDL; Cholesterol metabolism; Repeat; Signal.  
 FT SIGNAL 1 18 BY SIMILARITY.  
 FT PROPEP 19 24 BY SIMILARITY.  
 FT CHAIN 25 264 APOLOPROTEIN A-I.  
 FT DOMAIN 67 264 10 X APPROXIMATE TANDEM REPEATS.  
 FT REPEAT 67 88 1.  
 FT REPEAT 89 110 2.  
 FT REPEAT 111 121 3 (HAUF-LENGTH).  
 FT REPEAT 122 143 4.  
 FT REPEAT 144 165 5.  
 FT REPEAT 166 187 6.  
 FT REPEAT 188 209 7.  
 FT REPEAT 210 231 8.  
 FT REPEAT 232 242 9 (HAUF-LENGTH).  
 FT REPEAT 243 264 10.  
 SQ SEQUENCE 264 AA; 30659 MW; 7546F4BAF2846900 CRC64;

Query Match 9.1%; Score 171.5; DB 1; Length 264;  
 Best Local Similarity 21.6%; Pred. No. 0.00067;  
 Matches 72; Conservative 63; Mismatches 100; Indels 99; Gaps 11;  
 OY 7 VLTWALALLSFASTQARKGFWDYFSQTSQDGRVGEQIHQOKMARREPAITLKDSLEQ---- 62

DB 5 VVTALLFL---TGTQARY-FWQH-DEQAPLDLRLDYLVYLETVAASGDALAQFEAS 59  
 OY 63 -----DINNNKFLKXLRPLSGSEAPRLPODPV-----GMRKQLEHLE 101  
 DB 60 AVGKQLD---KLADNLDL-LGAAAKLREDMAPYKEVREMLKDTESLRAELTDLE 114  
 OY 102 EVKARLQPYMAEAEHLEIWMNEGLRQOLKPYTNLMEEVVALRYQELQRLVAGEEDTKAQ 161  
 DB 115 EVKERTIRPFLDQFSAKMTVEELOYRQRLAPAEELKLTOKVRLMOOKLTPVAEFAHRD 174  
 OY 162 LLGGVDEAMALLQGLQSRVYHHTGRFELFIPYAEISLVSGIRVQELHRSVAPAPASP 221  
 DB 175 LRG-----IYEELRKMLAPYSDELRL 194  
 OY 222 ARISRCYQVIVSRKLTAKALHARIQONLDOLKEHLSRAFACTGTEGACHPDQMLSEV 281  
 DB 195 QKLSQKLEIEIREKGIPOAEYQAKVVEQLSNLRKMT-----PLYODF 237  
 OY 282 RQRLQAFRODTYLOIAAFTRAIDQETEEVQOQ 315  
 DB 238 KERLTPTVAENLK-----TRFISL-LDRLOKTVVA 264

RESULT 15  
 APAL\_PIG STANDARD; PRT: 265 AA.  
 ID APAL\_PIG STANDARD; PRT: 265 AA.  
 AC P18648;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Apolipoprotein A-I precursor (Apo-AI).  
 GN APOA1.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.  
 OX NCBI\_TaxID=9623;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-93224154; PubMed-8468059;  
 RA Bitchbauer A., Knipping G., Juritsch H., Aschauer H., Zechner R.;  
 CC "Characterization of the apolipoprotein AI and CIII genes in the  
 CC domestic pig";  
 CC Genomics 15:643-652(1993).  
 CC  
 CC [2]  
 CC SEQUENCE FROM N.A.  
 RC TISSUE=Liver, and Brain;  
 RX MEDLINE-94125128; PubMed-8294940;  
 RA Moeckel B., Zinke H., Flach R., Weiss B., Weller-Guetliet H.,  
 RA Gassen H.;  
 CC "Expression of apolipoprotein A-I in porcine brain endothelium in  
 CC vitro";  
 CC J. Neurochem. 62:788-798(1994).  
 CC  
 CC [3]  
 CC SEQUENCE OF 34-265 FROM N.A.  
 RC TISSUE-Brain;  
 RX MEDLINE-90132667; PubMed-2105375;  
 RA Weller-Guetliet H., Sommerfeldt M., Papandrikopoulou A., Mischek U.,  
 RA Bonitz D., Frey A., Grupe M., Scheerer J., Gassen H.G.;  
 CC "Synthesis of apolipoprotein A-I in pig brain microvascular  
 CC endothelial cells";  
 CC J. Neurochem. 54:444-450(1990).  
 CC  
 CC [4]  
 CC SEQUENCE OF 105-265 FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE-93154581; PubMed-8428656;  
 RA Tieu V.N., Hasler-Rapacz J., Rapacz J., Black D.D.;  
 CC "Sequences and expression of the porcine apolipoprotein A-I and C-III  
 CC mRNAs";  
 CC Gene 123:173-179(1993).  
 CC  
 CC [5]  
 CC SEQUENCE OF 25-265.  
 RA Hasler-Rapacz J.O., Chaudhary R., Chowdhary B.P., Tieu V.N.,

RA Jackson K., Gustavsson I., Rapacz J.;  
 RI Submitted (Oct-1995) to the SWISS-PROT data bank.  
 RN [6]  
 RP MEDLINE-76184721; PubMed-178359;  
 RX Mahley R.W., Weisgraber K.H., Innerarity T., Brewer H.B. Jr.;  
 RT "Characterization of the plasma lipoproteins and apoproteins of the  
 RT Erythrocytes of the monkey."  
 RT Biochemistry 15:1928-1933(1976).  
 CC -1- FUNCTION: APOA-1 PARTICIPATES IN THE REVERSE TRANSPORT OF  
 CC CHOLESTEROL FROM TISSUES TO THE LIVER FOR EXCRETION BY PROMOTING  
 CC THE LIPIDATION OF CHOLESTEROL AND BY ACTING AS A COFACTOR FOR  
 CC THE LIPIDATION OF CHOLESTEROL AND BY ACTING AS A COFACTOR FOR  
 CC -1- TISSUE SPECIFICITY: MAJOR PROTEIN OF PLASMA HDL, ALSO FOUND IN  
 CC CHYLOMICRONS. SYNTHESIZED PREDOMINANTLY IN THE INTESTINE AND THE  
 CC LIVER.  
 CC -1- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: L00626; AAA30992.1; -  
 DR EMBL: X69477; CAA49234.1; -  
 DR EMBL: X17057; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL: X59414; CAA42050.1; -  
 DR PIR: J00704; J00704.  
 DR PIR: A05311; A05311.  
 DR PIR: A46018; A46018.  
 DR PIR: S21830; S21830.  
 DR PIR: S31394; S31394.  
 DR HSSP: P02647; IAVI.  
 DR InterPro: IPR000074; Apolipoprotein.  
 DR Pfam: PF01442; Apolipoprotein.1.  
 DR KJ Plasma: Lipid transport; HDL; Cholesterol metabolism; Repeat; Signal.  
 FT SIGNAL 1 18  
 FT PROPEP 19 24  
 FT CHAIN 25 265  
 FT DOMAIN 67 265  
 FT REPEAT 67 88  
 FT REPEAT 89 110  
 FT REPEAT 111 121  
 FT REPEAT 122 142  
 FT REPEAT 143 165  
 FT REPEAT 166 187  
 FT REPEAT 188 209  
 FT REPEAT 210 230  
 FT REPEAT 231 241  
 FT REPEAT 242 265  
 FT CONFLICT 108 108  
 FT CONFLICT 173 173  
 FT CONFLICT 173 143  
 FT CONFLICT 180 180  
 FT CONFLICT 185 186  
 FT CONFLICT 209 209  
 FT CONFLICT 224 224  
 FT CONFLICT 224 224  
 FT SEQUENCE 265 AA; 30325 MM; 2C6E578318ECP69C CRC64;  
 Query Match 9.1%; Score 170; DB 1; Length 265;  
 Best Local Similarity 22.3%; Pred. No. 0.00083;  
 Matches 73; Conservative 52; Mismatches 120; Indels 82; Gaps 8;

DB 60 ALGKHILMLKLDMDSLGTFKVRQGLGVNQGFMNDLKEFTEALRQENSKILFEVKKK 119  
 QY 107 LOPYMAEAEELVGMNLEGLRQOLKPYTMDLMEVVALRVQELQRLKVVGEOTQAQLLGCV 166  
 DB 120 VQYLDLDFONKMOEEMETVROKMAPLGAEREGARORVQELQKTL 164  
 QY 167 DEAMALLQGLQSVVHHTGCFKELFHPYAESIVSGICRHVQELHRSVAPHPAPASPARISK 226  
 DB 165 -----SPLAEELDRLEAHEALRQIVAPYSDDLRORMMA 199  
 QY 227 CVOVLSRKLTAKAKALHARIOQMLDQLRELSRAFGTGEAGDPQMLSEEVQRQLQ 286  
 DB 200 RPEALKEGCSLAE-VQAKAQEQKALGFKAKVLA-----KDLMOGLL 241  
 QY 287 AFRODTYLOIATFRAIDQTEVEVQOQ 313  
 DB 242 PVLENLKVSTLA--AIDEASKRLINQ 265

Search completed: January 23, 2003, 17:44:23  
 Job time: 21 secs

OY 5 AAVLFWALALISAFSATQAR-----KGFWDYFSQNSGDKR--VQIHOQ 47  
 DB 3 AAVLFWALALISAFSATQAR-----KGFWDYFSQNSGDKR--VQIHOQ 47  
 OY 48 KVARB-PATLKDSLQDQDNNMKNKFLKELRPLSGSEAPRLQDPVGMRRQLOEELVEYKAR 106



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: January 23, 2003, 17:36:06 : Search time 76 seconds  
(without alignments)  
992.280 Million cell updates/sec

Title: US-09-842-364-3  
Perfect score: 1877  
Sequence: 1 MASMAVLTMALALSAFSA.....LMEDITHSLHDGSHLGR 366

Scoring table: BIOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

SPREMBL\_21:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriophage:\*
- 17: sp\_archae:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARY

Result No.	Score	Query Match	Length	DB ID	Description
1	1864	99.3	363	4	Q9UBJ3
2	1310	69.8	368	11	Q9BP64
3	1301	69.3	368	11	Q9IX90
4	1293	68.9	367	11	Q9QDH3
5	316	16.8	366	13	Q93601
6	302.5	16.1	395	11	Q91XFB
7	299.5	16.0	395	11	Q9DBNO
8	289.5	15.4	435	11	Q91XFB
9	188	10.0	244	4	Q13784
10	172	9.2	264	11	Q922L4
11	165	8.8	990	4	Q9NYF9
12	164	8.7	317	6	Q9GLM7
13	163.5	8.7	1162	16	Q91316
14	163	8.7	259	13	Q98TGI
15	163	8.7	317	6	Q9GLM8
16	162	8.6	313	6	Q9GLC0

17	162	8.6	317	6	Q9GJ03	Q9GJ03 pan troglod
18	160	8.5	317	6	Q9GLM6	Q9GLM6 hylobates 1
19	159.5	8.5	275	13	Q9PT02	Q9PT02 oncorhynch
20	156.5	8.3	259	13	Q98TGS	Q98TGS anguilla ja
21	155.5	8.3	263	11	Q08855	Q08855 mus musculu
22	155.5	8.3	2310	5	Q9NKR1	Q9NKR1 leishmania
23	153.5	8.2	259	13	Q98TGS	Q98TGS anguilla ja
24	152.5	8.1	258	11	Q08877	Q08877 ratius norv
25	152.5	8.1	263	11	Q09042	Q09042 mus musculu
26	151.5	8.1	1936	5	Q20641	Q20641 caenorhabdi
27	151	8.0	241	6	Q9TS49	Q9TS49 erinaceus e
28	151	8.0	466	4	Q60374	Q60374 homo sapien
29	151	8.0	1313	4	Q75033	Q75033 homo sapien
30	151	8.0	2354	5	Q9NKT9	Q9NKT9 leishmania
31	150.5	8.0	591	5	Q17117	Q17117 caenorhabdi
32	149.5	8.0	258	11	Q09054	Q09054 ratius norv
33	147.5	7.9	1248	16	Q8U011	Q8U011 agrobacteri
34	147	7.8	800	3	Q96X03	Q96X03 emericeila
35	144.5	7.7	259	13	Q98TGS	Q98TGS anguilla ja
36	144.5	7.7	259	13	Q98TGS	Q98TGS anguilla ja
37	143.5	7.6	263	13	Q98TGS	Q98TGS anguilla ja
38	141.5	7.5	1000	2	Q54762	Q54762 synecococc
39	140	7.5	2779	5	Q9W4N7	Q9W4N7 drosophila
40	139.5	7.4	590	6	Q9BGY1	Q9BGY1 macaca fasc
41	139.5	7.4	806	11	Q8YD04	Q8YD04 mus musculu
42	139	7.4	174	13	Q98S13	Q98S13 cyprinus ca
43	138.5	7.4	552	5	Q967G6	Q967G6 trichinella
44	138.5	7.4	1208	4	Q9P2M7	Q9P2M7 homo sapien
45	137.5	7.3	1210	11	Q91VW5	Q91VW5 mus musculu

## ALIGNMENTS

## RESULT 1

Q9UBJ3 ID: Q9UBJ3 PRELIMINARY: PRT: 363 AA.

AC Q9UBJ3: 01-MAY-2000 (TREMURel. 13, Created)

DT 01-MAY-2000 (TREMURel. 13, Last sequence update)

DT 01-JUN-2001 (TREMURel. 17, Last annotation update)

DE Regeneration associated protein 3.

GN RAP3.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=LIVER.

RA Van der Vliet H.N., Keltzma P.H., Boers W., Chamuleau R.A.P.M.;

RT Identification of the human analog of rat RAP3.\*

RI Submitted (NOV-1999) to the EMBL/GenBank/DDJ databases.

DR EMBL: AF202890; AAF25662.1; -

DR EMBL: AF202899; AAF25661.1; -

DR InterPro: IPR000074; Apolipoprotein.

DR Pfam: pf01442; Apolipoprotein.

SQ SEQUENCE 363 AA; 40923 MW; 9E09CE97205D8A7D CRC64;

Query Match 99.3% Score 1864; DB 4; Length 363;  
Best Local Similarity 100.0% Pred. No. 1.8e-112;  
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	4	MAAVLTALALISAFSAQAARKGFWDFSGTSGDKGRVEQIHQOKMAREPATLKDLSLEQD	63
DB	1	MAAVLTALALISAFSAQAARKGFWDFSGTSGDKGRVEQIHQOKMAREPATLKDLSLEQD	60
QY	64	LNNNKKLEKLRPLSGSEAPRLPDDPYGMKROLELEEVYKARQIPMAAHELVGNLE	123
DB	61	LNNNKKLEKLRPLSGSEAPRLPDDPYGMKROLELEEVYKARQIPMAAHELVGNLE	120
QY	124	GLROOLPYTMDLMEQVALRQVLEQLOELRVYGEPTKQQLLGGVDEAVALLOGLSRVNH	183

Db 121 GLRQQLKPYTMDLMEQVALRVGELOQLRVGEDTKAQLLGGVDKAMALLQGLQSRVHH 180  
 QY 184 TGRFELFHPAASLVSGIGHVQVQLHRSVAPAPASPARLSRCVOVLSKRLTKAKALH 243  
 Db 181 TGRFELFHPAASLVSGIGHVQVQLHRSVAPAPASPARLSRCVOVLSKRLTKAKALH 240  
 QY 244 ARIQONLQDLREELSRFAGTGTGEGAGDPDMLEEVNQRLOAFRODTYLOIAAFTRAI 303  
 Db 241 ARIQONLQDLREELSRFAGTGTGEGAGDPDMLEEVNQRLOAFRODTYLOIAAFTRAI 300  
 QY 304 DQETEEVOOQLAPPPGHSAPAFEPQOTDSCKVLSKLQARLDLMDLMDITHSLDQSHL 363  
 Db 301 DQETEEVOOQLAPPPGHSAPAFEPQOTDSCKVLSKLQARLDLMDLMDITHSLDQSHL 360  
 QY 364 GDP 366  
 Db 361 GDP 363

## RESULT 2

099P64 PRELIMINARY: PRT: 368 AA.  
 AC 099P64  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Regeneration-associated protein 3.  
 CN APOA5 OR RAP3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN 111  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=C57BL/6J; TISSUE=LIVER.  
 RA Van der Vliet H.N., Groenink M., Leegwater A.C.J., Chamuleau R.A.F.M.;  
 RT "Identification of the mouse analog of rat RAP3."  
 RL Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF337059; AAC49600.1; -  
 DR MGD: MGI:1913363; APOA5.  
 DR InterPro: IPR000074; APOlipoprotein.  
 DR Pfam: PF01442; APOlipoprotein; 1.  
 SQ SEQUENCE 368 AA; 41320 MW; A55DF15BD0784696 CRC64;

Query Match 69.8%; Score 1310; DB 11; Length 368;  
 Best Local Similarity 72.6%; Pred. No. 8.7e-77;  
 Matches 267; Conservative 28; Mismatches 63; Indels 10; Gaps 5;

QY 4 MAAVLTWALALSAFSAATQARKGFWDYFSQTSQDKGRVQIHOQKMARPATLKDSLEOD 63  
 Db 1 MAAVLTWALALSAFSAATQARKGFWDYFSQTSQDKGRVQIHOQKMARPATLKDSLEOD 56  
 QY 64 LNNMKFLEKLRL--SGSEAPRLPDDPGMRKQLOEELLEVYKARLOPYMAEHLVGN 121  
 Db 57 LNNMKFLEKLRL--SGSEAPRLPDDPGMRKQLOEELLEVYKARLOPYMAEHLVGN 116  
 QY 122 LEGLRQQLKPYTMDLMEQVALRVGELOQLRVGEDTKAQLLGGVDKAMALLQGLQSRV 181  
 Db 117 LEGLRQQLKPYTMDLMEQVALRVGELOQLRVGEDTKAQLLGGVDKAMALLQGLQSRV 176  
 QY 182 HHTGFKELFHPAASLVSGIGHVQVQLHRSVAPAPASPARLSRCVOVLSKRLTKAKA 241  
 Db 177 HHTGFKELFHPAASLVSGIGHVQVQLHRSVAPAPASPARLSRCVOVLSKRLTKAKA 236  
 QY 242 LARIQONLQDLREELSRFAGTGTGEGAGDPDMLEEVNQRLOAFRODTYLOIAA 298  
 Db 237 LARIQONLQDLREELSRFAGTGTGEGAGDPDMLEEVNQRLOAFRODTYLOIAA 295  
 QY 299 FTRAIQDQETEEVOOQLAPPPGHSAPAFEPQOTDSCKVLSKLQARLDLMDLMDITHSLDQ 358  
 Db 296 FTRAIQDQETEEVOOQLAPPPGHSAPAFEPQOTDSCKVLSKLQARLDLMDLMDITHSLDQ 355  
 QY 359 GHSLLGDP 366

Db 356 GHSLLGDP 363

## RESULT 3

091X90 PRELIMINARY: PRT: 368 AA.  
 AC 091X90  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Unknown (protein for MGI:18812).  
 GN APOA5.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN 111  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=C57BL/6J; TISSUE=LIVER.  
 RA Strausberg R.;  
 RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: BC011198; AAH1198.1; -  
 DR MGD: MGI:1913363; APOA5.  
 DR InterPro: IPR000074; APOlipoprotein.  
 DR Pfam: PF01442; APOlipoprotein; 1.  
 SQ SEQUENCE 368 AA; 41237 MW; 174DB2481B88A660 CRC64;

Query Match 69.3%; Score 1301; DB 11; Length 368;  
 Best Local Similarity 72.0%; Pred. No. 3.3e-76;  
 Matches 265; Conservative 29; Mismatches 64; Indels 10; Gaps 5;

QY 4 MAAVLTWALALSAFSAATQARKGFWDYFSQTSQDKGRVQIHOQKMARPATLKDSLEOD 63  
 Db 1 MAAVLTWALALSAFSAATQARKGFWDYFSQTSQDKGRVQIHOQKMARPATLKDSLEOD 56  
 QY 64 LNNMKFLEKLRL--SGSEAPRLPDDPGMRKQLOEELLEVYKARLOPYMAEHLVGN 121  
 Db 57 LNNMKFLEKLRL--SGSEAPRLPDDPGMRKQLOEELLEVYKARLOPYMAEHLVGN 116  
 QY 122 LEGLRQQLKPYTMDLMEQVALRVGELOQLRVGEDTKAQLLGGVDKAMALLQGLQSRV 181  
 Db 117 LEGLRQQLKPYTMDLMEQVALRVGELOQLRVGEDTKAQLLGGVDKAMALLQGLQSRV 176  
 QY 182 HHTGFKELFHPAASLVSGIGHVQVQLHRSVAPAPASPARLSRCVOVLSKRLTKAKA 241  
 Db 177 HHTGFKELFHPAASLVSGIGHVQVQLHRSVAPAPASPARLSRCVOVLSKRLTKAKA 236  
 QY 242 LARIQONLQDLREELSRFAGTGTGEGAGDPDMLEEVNQRLOAFRODTYLOIAA 298  
 Db 237 LARIQONLQDLREELSRFAGTGTGEGAGDPDMLEEVNQRLOAFRODTYLOIAA 295  
 QY 299 FTRAIQDQETEEVOOQLAPPPGHSAPAFEPQOTDSCKVLSKLQARLDLMDLMDITHSLDQ 358  
 Db 296 FTRAIQDQETEEVOOQLAPPPGHSAPAFEPQOTDSCKVLSKLQARLDLMDLMDITHSLDQ 355  
 QY 359 GHSLLGDP 366  
 Db 356 GHSLLGDP 363

## RESULT 4

090UH3 PRELIMINARY: PRT: 367 AA.  
 AC 090UH3  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE Regeneration associated protein 3.  
 GN RAP3.  
 OS Rattus norvegicus (rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;  
 RN  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=WISTAR; TISSUE=LIVER;  
 RA Van der Vliet H.N., Groenink M., Leegwater A.C.J., Reitsma P.H.,  
 RA Boers W., Chamuleau R.A.F.M.;  
 RI \*isolation of a novel gene associated with liver regeneration.\*;  
 RI Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF202888; AAF25660.1;  
 DR EMBL: AF202887; AAF25659.1;  
 DR InterPro: IPR000074; Apolipoprotein.  
 DR Pfam: PF01442; Apolipoprotein; 1.  
 SQ SEQUENCE 367 AA; 41426 MW; D5B1A4F4C8016478 CRC64;

Query Match 68.9%; Score 1293; DB 11; Length 367;  
 Best Local Similarity 71.8%; Pred. No. 1,le-75;  
 Matches 260; Conservative 32; Mismatches 62; Indels 8; Gaps 4;

QY 4 MAAVITMALALLSAFSAATGARKGFMDYFSQTSQDKGRVQIHOQKMARPPATLKDSLKD 63  
 DB 1 MAAVITMALALLSVFATVQARKSFMEYFGNSQSGKMGQ--QOKLAE--SLKSLSED 56  
 QY 64 LNNMFKLEKRLPL--SSSEAPRLPDDPYGMRRQLOEELEEVKARLOPYMAEHLYGN 121  
 DB 57 LVNNMFKLEKRLPRLPKPEKPEPLADDEGIRKQLOEELEEVSTRLEPYMAKHQOVGN 116  
 QY 122 LEGLRQQLKPYTMDLMEQVALRVQELQELRVVGEEDTKAQLLGVDENALLQGLSRV 181  
 DB 117 LEGLRQQLKPYTVELMEQVGLSVQDLOELRVVKGKTKAQLLGVDENALLQDMSRYL 176  
 QY 182 HHTGRKELPHRYAESLVGICRHVOELHRSVAPAPASPARLSKCVLSKRLTKAKA 241  
 DB 177 HHTDRKELPHRYAESLVGICRHVOELHRSVAPAPASPARLSKCVLSKRLTKAKA 236  
 QY 242 LHRIGQNDQLEESR--AFAGTGEAGPDDPOMLSEEVKORLOARODTYLOIAF 299  
 DB 237 LHTSIQNDQLEESTIRVSTGADNRDSDQALSDVQRORLOARHDTYLOIAF 296  
 QY 300 TRAIQDETEVVOQLAPPPGSHSAPEFQGTDSKVLKLAARDLDMEDITHSLHDS 359  
 DB 297 TQAIQDETEIHOQLAPPPSHSAPELGHSDSKSLRLSLRDLDMEDIAVGLHDS 356  
 QY 360 HS 361  
 DB 357 HS 358

RESULT 5  
 O93601 PRELIMINARY; PRT; 366 AA.

AC O93601;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Apolipoprotein AIV.  
 GN APOAIV.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98221191; PubMed=9553114;  
 RA Steinmetz A., Hermann M., Nimpf J., Aebbersold R., Ducet A.,  
 RA Weinberg R.B., Schneider W.J.;  
 RT "Expression and conservation of apolipoprotein AIV in an avian  
 RT species.";  
 RL J. Biol. Chem. 273:10543-10549(1998).  
 DR EMBL: Y16534; CAA76273.1;  
 DR InterPro: IPR000074; Apolipoprotein.  
 DR Pfam: PF01442; Apolipoprotein; 2.  
 KW Lipoprotein.

SQ SEQUENCE 366 AA; 40852 MW; B47DB49F3E8D91CE CRC64;

Query Match 16.8%; Score 316; DB 13; Length 366;  
 Best Local Similarity 27.2%; Pred. No. 9,le-13;  
 Matches 91; Conservative 71; Mismatches 149; Indels 24; Gaps 7;

QY 1 MASMAVITMALALLSA--FSATGARKGFMDYFSQTSQDKGRVQIHOQKMARPPATLK 56  
 DB 1 MSPKALLVLLAVAGTRADVSPDQAVATLMRYFELGSSAKETVDQLOQAETKQNTL 60  
 QY 57 KSLSEODLNNMFKLEKRLPL--SSSEAPRLPDDPYGMRRQLOEELEEVKARLOPYMA 112  
 DB 61 ---IOSNANSYAEELRRLVPEATELOAQIYDPSOKLKQLOEELEEVKARLOPYMA 117  
 QY 113 EAHLEVGNNLEGLRQQLKPYTMDLMEQVALRVQELQELRVVGEEDTKAQLLGVDENALL 172  
 DB 118 EYHQDQIGTIRRELQAKLSYADELSQVDRGELRRALPEPATLRELDQMSIQAS 177  
 QY 173 LV---OGLSRVYVHHTGRKELPHRYAESLVGICRHVOELHRSVAPAPASPARLSKCV 228  
 DB 178 LGPYAERLQOQIDSSVEGLKGLTPLADELKEQVQSVGELKRLGSLPYAOEYVDGLNRL 237  
 QY 229 QVLSKRLTKAKALHARIQNDQLR-----EELSRFAG--TGTEGAGPDDPOMLSE 279  
 DB 238 QSLTQMERAEELNSKRLAASSEEMRAQSLAOLQALRDARAMQOLAPLAQILD 297  
 QY 280 EYVORLOAFRODTYLOIAFTRAIDETEVOQL 314  
 DB 298 RLQGYEARFQQAAPISFTFRQOLVRLKEMQKL 332

RESULT 6  
 O91XF8 PRELIMINARY; PRT; 395 AA.

AC O91XF8;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Unknown (protein for MGC:18592).  
 GN APOA4.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-KIDNEY;  
 RA Strausberg R.;  
 RI Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: M61:88051; APOA4.  
 DR InterPro: IPR000074; Apolipoprotein.  
 DR Pfam: PF01442; Apolipoprotein; 2.  
 SQ SEQUENCE 395 AA; 45029 MW; C4BBB32FMD41F71 CRC64;

Query Match 16.1%; Score 302.5; DB 11; Length 395;  
 Best Local Similarity 24.7%; Pred. No. 7,5e-12;  
 Matches 92; Conservative 84; Mismatches 164; Indels 33; Gaps 7;

QY 5 AAVITMALALLSA--FSATGARKGFMDYFSQTSQDKGRVQIHOQKMARPPATLK 58  
 DB 5 AAVITMALALLSVFATVQARKSFMEYFGNSQSGKMGQ--QOKLAE--SLKSLSED 64  
 QY 59 SLEQDNNMFKLEKRLPL--SSSEAPRLPDDPYGMRRQLOEELEEVKARLOPYMAEHLY 118  
 DB 65 KLGDASTVADGVNNKLVPRVVOISGLAQETEVKEIKLEEDLRDRMPHANKVDTGF 124  
 QY 119 GNNLEGLRQQLKPYTMDLMEQVALRVQELQELRVVGEEDTKAQLLGVD---EAMALIQ 174  
 DB 125 GENMOKLQELHRYAVNDLQDQINTQGEKRLQDTPIQNMOTTIENVDNHTSMPLAT 184  
 QY 175 GLQSRVYVHHTGRKELPHRYAESLVGICRHVOELHRSVAPAPASPARLSKCVLSRK 234

```

Db 185 NLKDFNRMMEELKGLTPRANELKATIDQNLDELRLSLAPLVGVOKLHNQMGSLAFQ 244
QY 235 LTLKAKALHARIQONLDOLRELSRAFACT-----GTBGAQDPDOMISEEVRQRLQAFR 289
Db 245 MKKNAEELQTKVSAKIDLOLKNLAPLVEDVQSKVKNTEGLOKSLDELINROEQVEEFR 304
QY 290 QDTYQIAAFRAIDQETEEVQOQLAPPPGHSAPAFPEQOTDSCKV---LSKIQARLDD 346
Db 305 KRVPMGEMFNKALVQQLQEQFRQQLCP-----NSGEVSHLSFLEKSLRE 349
QY 347 LMEJDTTHSLHDGC 359
Db 350 KVNSEFSTLEKKG 362

RESULT 7
Q9DBNO PRELIMINARY: PRT: 395 AA.
ID Q9DBNO:
AC Q9DBNO:
DT 01-JUN-2001 (TREMBLrel. 17, created)
DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE Adult male liver cDNA, RIKEN full-length enriched library,
DE clone:130002K10, full insert sequence.
GN APOA4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Aichi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,
RA Sato T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuhl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirni L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,
RA Blake J., Boilelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gusticich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
RA Lyons P., Marchionni I., Mashima J., Mazzarelli J., Momberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitlaker C., Wilmink L.,
RA Wyshaw-Borisi A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL: AK004856; BAB23620.1; -
DR MGD: MGI:88051; APOA4.
DR InterPro: IPR000074; Apolipoprotein.
DR Pfam: PF01442; Apolipoprotein.
SQ SEQUENCE 395 AA; 45044 MW; 4102D84ACBD0182A CRC64;

Query Match 16.0%; Score 299.5; DB 11; Length 395;
Best Local Similarity 24.7%; Pred. No. 1.2e-11;
Matches 92; Conservative 83; Mismatches 169; Indels 33; Gaps 7;

```

```

Db 125 GENMKLQENLKPVAVDLQDQINTQTEMKLQTLTYIQMOTIKENVDNLHTSMPLAT 184
QY 175 GLQSVNHHHTGFKELFPHYAESLVSGICRHVOELHRVAPAPASPARISFCVYLISK 234
Db 185 NLKDFNRMMEELKGLTPRANELKATIDQNLDELRLSLAPLVGVOKLHNQMGSLAFQ 244
QY 235 LTLKAKALHARIQONLDOLRELSRAFACT-----GTBGAQDPDOMISEEVRQRLQAFR 289
Db 245 MKKNAEELQTKVSAKIDLOLKNLAPLVEDVQSKVKNTEGLOKSLDELINROEQVEEFR 304
QY 290 QDTYQIAAFRAIDQETEEVQOQLAPPPGHSAPAFPEQOTDSCKV---LSKIQARLDD 346
Db 305 KRVPMGEMFNKALVQQLQEQFRQQLCP-----NSGEVSHLSFLEKSLRE 349
QY 347 LMEJDTTHSLHDGC 359
Db 350 KVNSEFSTLEKKG 362

RESULT 8
ID Q01488 PRELIMINARY: PRT: 435 AA.
AC Q01488:
DT 01-NOV-1996 (TREMBLrel. 01, created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE Variant apolipoprotein A-IV precursor (APOA-IV).
GN APOA-4.
OS Mus musculus castaneus (southeastern Asian house mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91286309; PubMed=1648102;
RA Reue K., Leete T.H.;
RT "Genetic variation in mouse apolipoprotein A-IV due to insertion and
RT deletion in a region of tandem repeats."
RL J. Biol. Chem. 266:12715-12721(1991).
CC -1- FUNCTION: MAY HAVE A ROLE IN CHYLOMICRONS AND VLDL SECRETION AND
CC CATABOLISM. REQUIRED FOR EFFICIENT ACTIVATION OF LIPOPROTEIN
CC LIPASE BY APOC-II; POTENT ACTIVATOR OF LCAT. APOA-IV IS A MAJOR
CC COMPONENT OF HDL AND CHYLOMICRONS. IT MAY PLAY A ROLE IN THE
CC INTRAASCULAR METABOLISM OF HDL. POTENTIAL ROLE IN CELLULAR
CC CHOLESTEROL EFFLUX.
CC -1- TISSUE SPECIFICITY: INTESTINE, LIVER, AND PLASMA.
CC (EACH 22-MER IS ACTUALLY A TANDEM ARRAY OF TWO, A & B, RELATED 11-
CC MEMERS) OCCURRING IN THIS SEQUENCE ARE PREDICTED TO BE HIGHLY ALPHA-
CC HELICAL, & MANY OF THESE HELICES ARE AMPHIPATHIC. THEY MAY
CC THEREFORE SERVE AS LIPID-BINDING DOMAINS WITH
CC LECTITHIN-CHOLESTEROL ACYLTRANSFERASE (LCAT) ACTIVATING ABILITIES.
CC -1- MISCELLANEOUS: THE APOA-IV POLYMORPHISMS ARE CLASSIFIED BASED ON
CC THE OCCURENCE OF THREE, FOUR, OR FIVE E-Q-A-Q/E-Q-O-Q/E-Q-V-Q
CC REPEAT UNITS.
DR EMBL: M64250; AAA37216.1; -
DR InterPro: IPR000074; Apolipoprotein.
DR Pfam: PF01442; Apolipoprotein.
KW Lipid transport; HDL; VLDL; Chylomicron; Repeat; Signal; Liver;
KW Plasma; Multigene family; Polymorphism; Cholesterol metabolism.
FT SIGNAL 1 20
FT CHAIN 1 435
FT DOMAIN 374 393
FT REPEAT 374 377
FT REPEAT 378 381
FT REPEAT 382 385
FT REPEAT 386 389
FT REPEAT 390 393
SQ SEQUENCE 435 AA; 49254 MW; B2AF55EAB8E456B2 CRC64;

Query Match 15.4%; Score 289.5; DB 11; Length 435;
Best Local Similarity 23.9%; Pred. No. 5.8e-11;
Matches 89; Conservative 85; Mismatches 166; Indels 33; Gaps 7;

```

```

QY 5 AAVLTWALLLSA----FSATQARKGFMDYFSQSGD-KGRVEIHQOKKAREPATL-KD 58
DB 5 AAVLTALVAITGTAKKVTSDVAVANVMDYFTHLSNNAKEAADQFOKTDYTOULSTLFD 64
QY 59 SLEODLNKKLEKLEKRLPSGSEAPRLPODPVGRROLOEELEEVKARLOPYMAEAHEL 118
DB 65 KLGDASTYADVGRNKLVPFVYQULSGHLAKETEVKEIKELLEDLDNRMPHANKVTQTF 124
QY 119 GWNLEGLROOLKPYTMDLMEQVALRVOELQOLRVNGEDTKAOLIGVD---EAMALLQ 174
DB 125 GEMMOKLEIKLKPYAVDLDOINTOTQEMKLOLTPYIQRQITIKENVDLHTSMPLAT 184
QY 175 GLOSRYVHHITGRKELFHPYAESLVSIGRHYOELHRSVAPHAPASPARLSRCVOYLSRK 234
DB 185 NIKDKFRNNKEELKGLHTTHANPLKATTDONLDRKSLAPITVGOVEKINHOMGLAFQ 244
QY 235 LTLKAKALHARIQONLDOLREELSRAFA9T---GTEGAGDPQOMLSEEVQRLQAFR 289
DB 245 MKKNABELQTKVSAKIDQLOKNLAPLVEDYQSKVKNTEGLQSKLDLNRQLODQVEEFR 304
QY 290 QDITYLOIAFATRAIDQETEVQOQIAPPPGHSAFAPEROOTISGRV---LSKIQARLD 346
DB 305 RTVEPGEEMFNKALVOLEQFROQLCP-----NSGEVESHLSLEKSLRE 349
QY 347 LMEDITHSLIDQ 359
DB 350 KVSFMSITLEKK 362

```

RESULT 9  
013784 PRELIMINARY: PRT: 244 AA.

```

ID 013784:
AC 013784:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE APOA4 protein (Fragment).
DE APOA4 protein (Fragment).
GN APOA4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_Taxid=9606;
OX NCBI_Taxid=9606;
RN 11
RP SEQUENCE FROM N.A.
RX TISSUE=LIVER;
RC MEDLINE=86016704; PubMed=3931073;
RA Karathanasis S.K.;
RT "Apolipoprotein multigene family: tandem organization of human
RT apolipoprotein A-I, CII, and AIV genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6374-6378(1985).
DR EMBL: M10373; AAB59516.1;
DR HSSP: P02649; 1B24.
DR InterPro: IPR000074; Apolipoprotein.
DR Pfam: PF01442; Apolipoprotein; 1.
FT NON_TER
SQ SEQUENCE 244 AA; 28157 MW; 01616ACB3F12734 CRC64;

```

Query Match 10.0%; Score 188; DB 4; Length 244;  
Best Local Similarity 26.1%; Pred. No. 9.7e-05;  
Matches 59; Conservative 44; Mismatches 97; Indels 26; Gaps 4;

```

QY 107 LQPYMAEAHELWGNLEGLRQOLKPYTMDLMEQVALRVOELQOLRVNGEDTKAOLIGV 166
DB 1 LQPYMAEAHELWGNLEGLRQOLKPYTMDLMEQVALRVOELQOLRVNGEDTKAOLIGV 166
QY 167 DEAMALLOGLQSRVYVHHTGRKELFHPYAESLVSIGRHYOELHRSVAPHAPASPARLSR 226
DB 61 EE-----LKGRLT-----PYADEFKVKIDQTEVEELRSLAPYAODTQEKLN 102
QY 227 CVOVLSRKLTLKAKALHARIQONLDOLREELSRAFA9T---GTEGAGDPQOMLSEEV 281
DB 103 QLEGTFQMKKNAEELKARISASAEELRORLAPLAEVGRNKLKNGTEGLQSLAEIGHL 162

```

```

QY 282 RORLQAFRODTYIQLAFAITRAIDQETEEVQOQLAP---PPPGSAF 324
DB 163 DQOVEFRFRREVPYGENFNKALVOQMEOLKQKLGPHAGDVEGHLSP 208

```

RESULT 10

```

ID 092214 PRELIMINARY: PRT: 264 AA.
AC 092214:
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Apolipoprotein A-I.
GN APOA1.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_Taxid=10036;
RN 11
RP SEQUENCE FROM N.A.
RX STRAIN=GOLDEN SYRIAN; TISSUE=INTESTINE;
RC MEDLINE=99061559; PubMed=9843713;
RA Wu J.Y., Reeves S.K., Wang Y.R., Wu Y., Lei P.P., Lei K.Y.;
RT "Zinc deficiency decreases plasma level and hepatic mRNA abundance of
RT apolipoprotein A-I in rats and hamsters.";
KL Am. J. Physiol. 275:C1516-C1525(1998).
DK EMBL: AF046919; AAC98484.1;
DR HSSP: P02647; IAV1.
DR InterPro: IPR000074; Apolipoprotein.
DR Pfam: PF01442; Apolipoprotein; 1.
KW Lipoprotein.
SQ SEQUENCE 264 AA; 30739 MW; 280B22F4C0B0129 CRC64;

```

Query Match 9.2%; Score 172; DB 11; Length 264;  
Best Local Similarity 21.8%; Pred. No. 0.0011;  
Matches 61; Conservative 57; Mismatches 112; Indels 50; Gaps 8;

```

QY 9 TMAIALISAF-SATQAR-----KGFMDYFSQSGDKR--VBOJHOKKA 50
DB 3 TVLAVALVLELTQSQAHEFWQDDPQTPMDRVKDFATVYVDAAKDSGREVYSOFETSLG 62
QY 51 RE-PATLKDSLEODLNKKLEKRLPSGSEAPRLPODPVGRROLOEELEEVKARLOP 109
DB 63 KQNLINLLEWMDLPGSVGVLQOLQPVTOFMDNLKETEEMLRRENNKLEEVKAYOP 122
QY 110 YMAEAHELWGNLEGLRQOLKPYTMDLMEQVALRVOELQOLRVNGEDTKAOLIGVD 167
DB 123 YLDQFO--TKWQEEVALYRQKMEPLGAEELRDGAKRQKLQELQELTLPDGD----- 170
QY 168 KAMALLQGLQSRVYVHHTGRKELFHPYAESLVSIGRHYOEL-----HRSVAPHAP 218
DB 171 -----LDRRHRIHVDALRTKMTPTSDQMDRLERLQDLSDPTLAEYHTRKADHKK 222
QY 219 ASPARLSRCVOYLSRKLTLKAKALHARIQONLDOLREELS 258
DB 223 AFGFKAKPALEDLRLQGLMPFESFTRIMSMVDEASKLN 262

```

RESULT 11

```

ID 09NYF9 PRELIMINARY: PRT: 990 AA.
AC 09NYF9:
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE Golgi matrix protein GML30.
GN GOLG32.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;

```

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-PLACENTA:  
 RA Weide T.B., Bayer M., Barnekow A.:  
 RT "Human g91g1 matrix protein 130, hucM130, a protein highly homologous  
 to GM130 from rat."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF248953; AAF65550.1;  
 KW Matrix protein.  
 SQ SEQUENCE 990 AA; 11658 MW; 2380FBED11B8DAF5 CRC64;

Query Match 8.8%; Score 165; DB 4; Length 990;  
 Best Local Similarity 22.0%; Pred. No. 0.017;  
 Matches 91; Conservative 63; Mismatches 128; Indels 132; Caps 16;

QY 15 LSASFATARKGFMD-----YFSQTSQDKGRVEQIHQOKMAREPATLKDSL 60  
 DB 260 LSASFATARKGFMD-----YFSQTSQDKGRVEQIHQOKMAREPATLKDSL 319  
 QY 61 EODLNMMKFLKRLPL-----SCSEAPRLPODPVGMRRQIQEELFEVKARIQPYMAEAH 115  
 DB 320 QLNLEELQKLEMTLELLQGFSSRCEAFDANO-----QLOQAMEF-RAQLEAHLGGVM 371  
 QY 116 ELVGNLEGLRQOLKPYMDIMEQVAL---RVQLEQELRVVGEDTKAKOLLGVDENAMAL 172  
 DB 372 E-----SVRLQEMENDKVAENLKGSAMWRKQMOQSEQVHTLRKREKCSM----- 417  
 QY 173 LOGIOSRVVHTGKRELFHPYAESLVSGIRHVOELHRSVA-----PHAPASP 221  
 DB 418 -----SRVOELSTSLAELRNMAEPPPPPPGAP 446  
 QY 222 ARLSKCVQ---LSRKLTLAKALHANIQON-----IDQRE---ELSRFACTGTEE 268  
 DB 447 SEVEOLOAEAEHLRKELEGLAGLOAVODNEGLSRLEQCEELTLERAAELMGQ- 505  
 QY 269 GAGDPQOMLSEVGRQIRQAFQDPTVQLAFAFTRAIQDTEVVOQLAPPPCHSAFAPEF 328  
 DB 506 -----AARROILTTMONDR---TTISRALSO-NREIKEDQALAELOSGPVKLINEN 551  
 QY 329 QQRTDS-----GKVLSTKLOARLDLMDI-----THSLHDQCHSLG 364  
 DB 552 MEITSALQSEQHVKEIKGKIGLEQEKISLKEIVELKSGVAGSIIQQQRDQYIC 605

RESULT 12  
 QY 09GLM7 PRELIMINARY; PRT; 317 AA.  
 AC 09GLM7;  
 DT 01-MAR-2001 (TREMBlrel, 16, Created)  
 DT 01-MAR-2001 (TREMBlrel, 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel, 19, Last annotation update)  
 DE Apolipoprotein E.  
 GN APOE.  
 OS Pongo pygmaeus (Orangutan).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pongo.  
 OX NCBI\_TaxID=9600;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kogaev E.I., Dvorianchikov G.A., Riazanskaya N.N.:  
 RT "APOE gene evolution in Homioidae."  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF200505; AAG28580.1;  
 DR EMBL: AF200503; AAG28580.1; JOINED.  
 DR EMBL: AF200504; AAG28580.1; JOINED.  
 DR HSP; P02649; IOEF.  
 DR InterPro: IPR000074; Apolipoprotein.  
 DR Pfam: PF01442; Apolipoprotein; 1.  
 KW Lipoprotein.  
 SQ SEQUENCE 317 AA; 36094 MW; 6B08CEC5B033BC2D CRC64;

Query Match 8.7%; Score 164; DB 6; Length 317;  
 Best Local Similarity 21.4%; Pred. No. 0.0047;

Matches 67; Conservative 58; Mismatches 120; Indels 68; Caps 8;

QY 10 WALALLSASFATQARKGFMDY---FSQTSQDKGRVEQIHQOKMAREPATLKDSL 67  
 DB 44 WELAL-----GFWMDYLRWQUTLSEQVQ--EELLSQVOTQELTALMDTMEKELKAY 92  
 QY 68 NKFL-EKLRPLSGSAPRLPODPVGMRRQIQEELFEVKARIQPYMAEAHVLGNMLDELRL 126  
 DB 93 KSELEQLPYAEETRALSKSELQAAQARLGADMDVGRVQYRGEVQAMGOSTEELR 152  
 QY 127 QOLKEPYTMDLMEQVALRVQLEQELRVVGEDTKAKOLLGVDENAMALLOGIOSRVVHTGK 186  
 DB 153 ARLASHLKRKLRLRLDADDDQKRLAVYQACARQAEKGV-----SALIKERLIGP 201  
 QY 187 FKELFHPYAESLVSGIRHVOELHRSVAPIAPASPARLSRCVQVLSRKLTLAKALHARI 246  
 DB 202 LVEQGRVRAATVGVSAKPIQER-----AQAMGERLRAEMEGSRT 243  
 QY 247 QONILQLEELSRFACTGTERGACPDHOMLSEVGRQIRQAFQDPTVQLAFAFTRAI--- 303  
 DB 244 RDRIDEVEQVA-----EVRKLEQADQOIRQAEFAQARLKS 282  
 QY 304 -DQTEVEVQOOLA 315  
 DB 283 FEPLVEDMQOWA 295  
 RESULT 13  
 QY 09I316 PRELIMINARY; PRT; 1162 AA.  
 AC 09I316;  
 DT 01-MAR-2001 (TREMBlrel, 16, Created)  
 DT 01-MAR-2001 (TREMBlrel, 16, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel, 21, Last annotation update)  
 DE Hypothetical protein PA1527.  
 GN PA1527.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-ATCC 15692 / PA01;  
 RA MEDLINE-20437337; PubMed-10984043;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Keizer J., Sailer M.H., Hancock R.F.W., Lory S., Olson M.V.:  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 opportunistic pathogen."  
 RL Nature 406:959-964(2000).  
 DR EMBL: AE004581; AAC04916.1;  
 DR InterPro: IPR003439; ABC transport.  
 DR InterPro: IPR003405; SMC\_C.  
 DR InterPro: IPR003395; SMC\_N.  
 DR Pfam: PF02483; SMC\_C; 1.  
 DR Pfam: PF02463; SMC\_N; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 1162 AA; 131580 MW; A62DDCC73015ACB5 CRC64;

Query Match 8.7%; Score 163.5; DB 16; Length 1162;  
 Best Local Similarity 21.8%; Pred. No. 0.026;  
 Matches 105; Conservative 59; Mismatches 153; Indels 165; Caps 20;

QY 3 SMAALVFWALLLSASFATQARKGFMDYFSQTSQDKGRVEQIHQOKMAREPATLKDSL 62  
 DB 359 SAAALAEKGLILDEQAEQGMQAMQOMDAFNQSAEPKQAEVQOSRI-----QHLEQ 410  
 QY 63 DLNNNNKFLKRLPLSGSEAPRLPODPVGM-----RQLOEE-----DEEV 103  
 DB 411 SLERIQDRERIRQERQGLAAD-PEDAAILELNQVAIAELALELQLOEGQAEERLEQL 469

